

Institut für Didaktik und Ausbildungsforschung in der Medizin  
der Ludwig-Maximilians-Universität München  
Vorstand: Univ.-Prof. Dr. med. Martin R. Fischer, MME (Bern)

**Messung von klinischen Diagnosekompetenzen  
bei Medizinstudierenden  
in einer simulierten Lernumgebung**

**Dissertation**

zum Erwerb des Doktorgrades der Medizin  
an der Medizinischen Fakultät der  
Ludwig-Maximilians-Universität zu München

vorgelegt von  
Victoria Maria Sophia Reitmeier

aus  
Erlangen

2020

Mit Genehmigung der Medizinischen Fakultät der Universität München

Berichterstatter: Prof. Dr. Martin Fischer

Mitberichterstatter: Prof. Dr. Ralf Schmidmaier

Prof. Dr. Martin Dreyling

Mitbetreuung: Dr. med. Hannah Gerstenkorn

Dekan: Prof. Dr. med. dent. Reinhard Hickel

Tag der mündlichen Prüfung: 26.05.2020

# Inhaltsverzeichnis

<b>Abkürzungsverzeichnis.....</b>	<b>V</b>
<b>Zusammenfassung.....</b>	<b>VI</b>
<b>1 Einleitung.....</b>	<b>1</b>
1.1 Hintergrund der Arbeit.....	1
1.2 Theoretische Grundlagen.....	5
1.2.1 Begriffsklärungen.....	5
1.2.2 Professionelle Wissensbasis.....	7
1.3 Diagnosekompetenzen.....	8
1.3.1 Entwicklung von Diagnosekompetenzen.....	8
1.3.2 Anwendung von Diagnosekompetenzen.....	9
1.3.3 Einflussfaktoren auf Diagnosekompetenzen.....	11
1.3.4 Lernumgebungen für Diagnosekompetenzen.....	13
1.3.5 Arbeitsmodell für Diagnosekompetenzen.....	15
1.3.6 Messinstrumente für Diagnosekompetenzen.....	16
1.4 Ziel der Arbeit und Fragestellungen.....	19
<b>2 Methodik.....</b>	<b>20</b>
2.1 Studiendesign.....	20
2.2 Studienteilnehmende.....	21
2.3 Studienmaterial.....	23
2.3.1 Inhalt des Studienmaterials.....	24
2.3.1.1 Konzeptueller Wissenstest.....	24
2.3.1.2 Key Feature Test.....	24
2.3.1.3 Simulationen.....	24
2.3.1.4 Pädagogisch-psychologische Messungen.....	26
2.3.2 Kodierung der Messinstrumente.....	26
2.3.2.1 Konzeptueller Wissenstest.....	26
2.3.2.2 Key Feature Test.....	26
2.3.2.3 Simulationen.....	26
2.4 Studiendurchführung.....	27
2.5 Statistische Analysen.....	29
2.5.1 Deskriptive Auswertung der Messinstrumente.....	29
2.5.2 Auswertung der Fragestellungen.....	30

<b>3</b>	<b>Ergebnisse .....</b>	<b>31</b>
3.1	Deskriptive Auswertung der Messinstrumente .....	31
3.1.1	Item- und Skalenanalyse der Wissenstests .....	31
3.1.2	Testergebnisse der Messinstrumente .....	35
3.2	Auswertung der Fragestellungen .....	48
3.2.1	Korrelation Performanz in Wissenstests .....	48
3.2.2	Korrelation Performanz in Wissenstests und Simulationen .....	48
3.2.3	Assoziationsfaktoren von Performanz in Simulationen .....	52
<b>4</b>	<b>Diskussion .....</b>	<b>55</b>
4.1	Zusammenfassung der Ergebnisse .....	55
4.1.1	Zusammenfassung der deskriptiven Analysen .....	55
4.1.1.1	Item- und Skalenanalyse der Wissenstests .....	55
4.1.1.2	Testergebnisse der Messinstrumente .....	55
4.1.2	Zusammenfassung der Fragestellungen .....	56
4.2	Diskussion der Ergebnisse .....	57
4.2.1	Diskussion der deskriptiven Analysen .....	57
4.2.1.1	Item- und Skalenanalyse der Wissenstests .....	57
4.2.1.2	Testergebnisse der Messinstrumente .....	58
4.2.2	Diskussion der Fragestellungen .....	59
4.3	Diskussion des Studienmaterials und der Methoden .....	62
4.4	Implikationen und zukünftige Forschungen .....	65
<b>5</b>	<b>Anhang .....</b>	<b>68</b>
<b>6</b>	<b>Verzeichnisse .....</b>	<b>85</b>
6.1	Literaturverzeichnis .....	85
6.2	Tabellenverzeichnis .....	92
6.3	Abbildungsverzeichnis .....	94
<b>7</b>	<b>Danksagung .....</b>	<b>95</b>
<b>8</b>	<b>Eidesstattliche Versicherung .....</b>	<b>96</b>

## Abkürzungsverzeichnis

AP	Angina Pectoris
BSG	Blutsenkungsgeschwindigkeit
CAP	Ambulant erworbene Pneumonie
CLEOS	Clinical Expert Operating System
CRP	C-reaktives Protein
DAM	Institut für Didaktik und Ausbildungsforschung in der Medizin am Klinikum der Ludwig-Maximilians-Universität München
DFG	Deutsche Forschungsgemeinschaft
DKMP	Dilatative Kardiomyopathie
EKG	Elektrokardiogramm
LMU	Ludwig-Maximilians-Universität München
LSB	Linksschenkelblock
KF	Key Feature
KFQ	Key Feature Frage
KW	Konzeptuelles Wissen
KWQ	Konzeptuelles Wissen Frage
MC	Multiple Choice
O <sub>2</sub>	Sauerstoff
OSCE	Objective Structured Clinical Examination
P	Studierende im Praktischen Jahr
RGs	Rasselgeräusche
RSB	Rechtsschenkelblock
S	Studierende im Klinischen Studienabschnitt
SpO <sub>2</sub>	Pulsoxymetrisch gemessene Sauerstoffsättigung
StaBLab	Statistisches Beratungslabor
TU	Technische Universität München
Z.n.	Zustand nach
ZeUS	Zentrum für Unterricht und Studium

## Zusammenfassung

### Einleitung

Diagnostizieren zählt zu den kritischsten Aufgaben eines Arztes und ist ein zentrales Messinstrument seiner Fähigkeiten. Das Anamnesegespräch ist eine der wichtigsten Möglichkeiten zweckgerichtet Patientendaten zu sammeln. Lernen anhand realitätsgetreuer Problemstellungen fördert Diagnosekompetenzen<sup>1</sup> und ist deshalb umfassend in die Ausbildung von Medizinstudierenden zu integrieren. Ziel der vorliegenden Arbeit ist, Instrumente zur Messung von klinischen Diagnosekompetenzen bei Medizinstudierenden in einer simulierten Lernumgebung zu evaluieren.

### Methode

Als Messinstrumente wurden 2 Wissenstests, Konzeptueller Wissenstest und Key Feature Test, sowie Patientenfälle für Anamnesegespräche zur Inhaltsdomäne Atemnot konzipiert. Anamnesegespräche wurden in Live- und in Video-Simulation durchgeführt; die Zuteilung der Reihenfolge erfolgte im Voraus. Die Testergebnisse wurden über die Online-Plattform CASUS erhoben, nach Studiumsstatus kategorisiert und von 47 Studierenden im Klinischen Studienabschnitt und 39 Studierenden im Praktischen Jahr ausgewertet. Die statistischen Analysen umfassten eine deskriptive Auswertung sowie quantitative Modelle zur Untersuchung von Zusammenhängen zwischen den Messinstrumenten.

### Ergebnisse

Im Konzeptuellen Wissenstest sowie im Key Feature Test war das durchschnittliche Testergebnis jeweils 66 %. Die Leistungen in den Simulationen waren stark fallbezogen bei einem durchschnittlichen Testergebnis von 40 %. Studierende im Praktischen Jahr erzielten bei allen Messinstrumenten bessere Testergebnisse. Die Performanz in den beiden Wissenstests korrelierte signifikant positiv ( $r = .73$ ;  $p < .01$ ). Korrelationen zwischen der Leistung in Wissenstests und Simulationen waren fallspezifisch, abhängig vom Studiumsstatus, unabhängig der Simulationsart. Live-Simulationen für Anamnesegespräche führten zu besseren Testergebnissen als Video-Simulationen. Konsistente Assoziationen zwischen einer Reihenfolge der Simulationsart und der Performanz bei der Durchführung von Anamnesegesprächen konnten nicht festgestellt werden.

---

<sup>1</sup> Boshuizen, H. P. and Schmidt, H. G., "The Development of Clinical Reasoning Expertise," in *Clinical Reasoning in the Health Professions*, 4th ed., J. Higgs, G. Jensen, S. Loftus, and N. Christensen, Eds. Edinburgh: Elsevier Health Sciences, 2019, pp. 57–66.

## **Diskussion**

Diagnosekompetenzen zeigten sich als nicht generalisierbare Kompetenzen, positiv durch medizinische Expertise beeinflusst. Live-Simulationen führten zu mehr akkuraten Diagnosen und sind deshalb weiterhin für die Erhebung der Anamnese in der Klinik zu empfehlen. Video-Simulationen bieten zukunftssträchtige Chancen für die Ausbildung und Messung von klinischen Diagnosekompetenzen bei Medizinstudierenden. Für statistisch gesicherte Empfehlungen bedarf es wiederholter Messungen im Rahmen kommender Studienphasen. Hierfür gilt es, inhaltliche und strukturelle Adaptierungen der Messinstrumente vorzunehmen.

# 1 Einleitung

## 1.1 Hintergrund der Arbeit

Diagnostizieren zählt zu den kritischsten Aufgaben eines Arztes. Als Hauptbestandteil seines beruflichen Handelns ist es eines der wichtigsten Messinstrumente seiner Fähigkeiten [1]. Richtig gestellte Diagnosen sind Grundlage für die adäquate Therapie einer Krankheit.

Irren ist menschlich, wie Seneca bereits feststellte [2], und führt im Rahmen des Diagnostizierens zu Fehldiagnosen. Die negativen Konsequenzen daraus betreffen in der Medizin Patienten, Ärzte und letztlich auch die Volkswirtschaft. Der Anteil an Fehldiagnosen beträgt circa 5 % in bildgebenden Fachrichtungen, erreicht in anderen Teilgebieten jedoch bis zu 15 % [3]. Fehldiagnosen sind zu spät gestellte, falsche oder fehlende Diagnosen, bei denen verfügbare Fakten inadäquat oder nicht rechtzeitig interpretiert wurden [4].

Ursachen von Fehldiagnosen liegen in System Errors, Cognitive Errors oder Non-fault Errors [4]. System Errors ergeben sich aus unzureichenden technischen oder apparativen Ausstattungen sowie Organisationsfehlern.

Wissenslücken beziehungsweise schlechte Datenerhebung oder -verknüpfung beim Diagnostiker führen zu Cognitive Errors. Daraus resultieren falsche Verdachtshypothesen, die im weiteren Diagnoseprozess fehlerhaft umformuliert und weiterentwickelt werden [5]. Heuristiken über den diagnostischen Wert klinischer Informationen werden verzerrt ausgelegt [6], die Wahrscheinlichkeit des Auftretens seltener Erkrankungen wird überschätzt, während die häufiger Erkrankungen unterschätzt wird [5].

Non-fault Errors entstehen aufgrund externer Faktoren, deren Ursachen nicht vom diagnostizierenden Arzt abhängig sind. Hierzu zählen untypische Formen eines Krankheitsbildes oder mangelnde Kooperation des Patienten.

Zur Entstehung von Fehldiagnosen trägt in den meisten Fällen jedoch nicht eine Fehlerquelle alleine, sondern die Kombination aus systembezogen und kognitiven Faktoren bei [4].

Fehldiagnosen haben nicht nur direkte und indirekte Auswirkungen auf die Genesung und das Wohl des Patienten, sondern sind ein hoher finanzieller Belastungsfaktor für das Gesundheitssystem. Der Schaden für den Patienten ergibt sich aus den negativen Konsequenzen von Untersuchungen, Interventionen und pharmakologischen Behandlungen, volkswirtschaftlicher Schaden entsteht aus den Kosten derselben. Konsequenzen für den Arzt zeigen sich in der Patientenzufriedenheit, der interkollegialen Kommunikation, im Zeitmanagement sowie in finanziellen und juristischen Folgen.



Die Kosten um Diagnosen zu testen stellen den am stärksten wachsenden Ausgabenfaktor dar [7]. Deshalb muss aus ökonomischer Sicht der Erwerb ärztlicher Diagnosekompetenzen bestmöglich gefördert werden, um die steigenden Ausgaben eindämmen zu können. Da die fehlerhafte Synthese verfügbarer Informationen Hauptursache für Cognitive Errors ist [4], gilt es nicht nur das Fachwissen zu verbessern, sondern auch die Fertigkeit des Diagnostizierens selbst zu trainieren. Um Fehler in der späteren ärztlichen Praxis zu vermeiden, sollte Diagnostizieren deshalb bereits im Studium trainiert und als eine elementare Komponente im medizinischen Curriculum verankert werden.

Das Anamnesegegespräch ist eine zentrale Diagnosesituation [8] und eine wichtige Möglichkeit, zweckgerichtet Daten zu sammeln. Etymologisch bedeutet Anamnese „sich erinnern“, wird aus dem Griechischen *ἀνάμνησις* abgeleitet. Aufgabe des Arztes bei der Erhebung der Diagnose ist es, dem Patienten zu helfen, sich an alle für den aktuellen Beratungsanlass relevanten Fakten zu erinnern. Mittels gezielter Fragestellungen des Diagnostikers werden neben relevanten symptombezogenen Informationen durch die biographische Anamnese, Familien- und Sozialanamnese, Medikamentenanamnese sowie vegetative Anamnese persönliche und psychosoziale Informationen erfasst. Das Anamnesegegespräch steht am Beginn des Arzt-Patientenkontakts und ist somit ein zentraler Baustein für den Aufbau der Arzt-Patientenbeziehung. Lediglich in besonderen Fällen wie bei bewusstlosen, dementen oder sehr jungen Patienten wird die Eigenanamnese durch eine Fremdanamnese ersetzt. Durch das Anamnesegegespräch werden bis zu 80 % aller für die Diagnose notwendigen Daten gesammelt, wodurch alleine bis zu 76 % aller Diagnosen gestellt werden können [8]. Die Durchführung des Anamnesegegesprächs ist kontext- und situationsabhängig, der inhaltliche Umfang und Fokus fachspezifisch stark variabel. Sowohl Arzt als auch Patient können entscheidend zur Ergebnisqualität beitragen. Die Dokumentation der Informationen erfolgt in einem Anamnesebogen, welcher als fester Bestandteil der Patientenakte neben medizinischer auch forensische Relevanz besitzt.

Anamnestizieren bedeutet jedoch nicht nur Daten zu sammeln, sondern auch auf die Emotionen des Patienten zu reagieren, diesen zu führen und auf dessen Verhalten einzuwirken [8]. Klinisch tätige Ärzte führen im Laufe ihres Berufslebens zwischen 100.000 und 200.000 Anamnesegegespräche [8]. Das Anamnesegegespräch stellt heterogene Herausforderungen an inhaltliche-, zwischenmenschliche-, und Kommunikationskompetenzen des Diagnostikers [8]. Diese Kompetenzen können durch anweisungsorientierte (z.B. Skripte, Online-Kurse) und erfahrungsorientierte (z.B. Rollenspiele) pädagogische Instruktionen trainiert werden [8].

Die zunehmenden technischen Automatisierungsprozesse der Medizin betreffen auch das Anamnestizieren. Erste computergestützte Anamnesesysteme sind bereits seit den 1960er Jahren verfügbar [9]. Das Clinical Expert Operating System (CLEOS) ist Beispiel für ein webbasiertes Anwendungssystem zur standardisierten Erfassung von Patientendaten und wird als möglicher Ersatz für das klassische Anamnesegegespräch geprüft [10]. Vorteile liegen in der anonymisierten standardisierten Erhebung, Verwaltung und Zugänglichkeit der Daten. Ein großer Nachteil besteht jedoch darin, dass nonverbale Kommunikation, als zentraler Bestandteil einer Anamnese, von computergestützten Systemen nicht erfasst wird. Diese ist jedoch oftmals entscheidend in einer Arzt-Patientenbeziehung, da die im Rahmen der nonverbalen Kommunikation wahrgenommenen Emotionen den weiteren Diagnose- und Therapieprozess stark beeinflussen können. Auch kontextspezifische Anpassungen an kognitive oder sprachliche Kompetenzen des Patienten sowie Patientensetting können bei der Computererfassung nicht berücksichtigt werden und deshalb zu verzerrten Ergebnissen führen. Die wissenschaftlichen Evidenzen hinsichtlich computergestützter Anamnesesysteme sind rar; Publikationen über den Einsatz von CLEOS in der medizinischen Praxis suggerieren jedoch, trotz der guten Akzeptanz bei Patienten, dass derartig computergestützte Systeme mehr für einen Expertiseaustausch unter Ärzten als für die Erfassung von Patienteninformationen und somit als Anamnesetool verwendet werden sollten [11].

Die Ausbildung von Medizinstudierenden ist an vielen Fakultäten immer noch stark auf den Wissenserwerb aus Vorlesungen und Sachbüchern fokussiert [12] und es werden wenige Gelegenheiten angeboten, theoretisches Wissen praktisch zu trainieren. Dieser fehlende Bezug zwischen akademischen und klinischen Lernumgebungen wird mittlerweile als eine Priorität bei der Überarbeitung der Lehrpläne angesehen [13]. Unter der Annahme, dass Anamnestizieren und Diagnostizieren im Laufe der Zeit von erfahrenen Kollegen erlernt werden könnten [14], wird die Unterrichtung dieser Kompetenzen in den Lehrplänen oft nicht spezifisch berücksichtigt. Weitere Gründe sind Kosten und Organisationsaufwand, welche sich durch die Implementierung zusätzlicher Lehrveranstaltungen in ein bestehendes Curriculum ergeben. Eigenes Handeln ist jedoch zentral für den Erwerb von Kompetenzen [15], denn häufiges Trainieren medizinischer Problemstellungen bildet umfangreiche Wissensnetzwerke für Krankheitsbilder [16]. Durch Praxis können theoretisches Wissen und erlernte Fertigkeiten verknüpft werden [17] und durch regelmäßiges Üben Kompetenzen im Anamnesegegespräch gefördert werden [18]. Darüber hinaus wird dadurch dem Entstehen von tragem Wissen, der Unfähigkeit erworbenes Wissen auf verschiedene Situationen zu übertragen, vorgebeugt [19]. Kompetenzen werden durch Ler-

nen anhand von realitätsgetreuen Patientenfällen in Kombination mit instruktionaler Unterstützung und Reflexionsphasen erworben und gefördert [17]. Ziel ist es, die kognitive Flexibilität der Studierenden zu maximieren. Durch verschiedene Lernumgebungen und Anwendungssituationen sollen domänspezifisches Wissen vernetzt, Wissenstransfer gefördert und somit flexibel anwendbare Kompetenzen erlernt werden [20].

Idealer Startzeitpunkt, diese Qualifikationen zu trainieren, ist der klinische Studienabschnitt, da zur Vermeidung von Cognitive Overload und somit für ein erfolgsversprechendes Trainieren medizinische Expertise erforderlich ist [21], [22] (s. 1.3.1). In Deutschland sind zur Entwicklung und Förderung von Anamnesekompetenzen verschiedene Lehr- und Praxisveranstaltungen im Curriculum etabliert: Blockpraktika, Famulaturen und das Praktische Jahr dienen unter anderem dazu, klinisch das Anamnestizieren zu trainieren [23]. Spezifische Kurse und Seminare zur ärztlichen Gesprächsführung unterliegen der Studienordnung der jeweiligen Fakultäten, sind jedoch nicht als Pflichtmodule in der Approbationsordnung erfasst [23]. Darüber hinaus bieten Universitäten die Möglichkeit, im Rahmen von Anamnesegruppen fakultativ zusätzliche Kompetenzen in der Patientenkommunikation zu trainieren [24]. Diese sind meist in der Form des Peer-Learnings, einem gemeinsamen Lernen von und mit Kommilitonen ohne professionelle Dozenten, organisiert. Ziel der Anamnesegruppen ist es, die Kommunikationstechniken zu verbessern sowie die Gestaltung der Arzt-Patienten-Beziehung und die Auseinandersetzung mit der künftigen ärztlichen Rolle zu fördern [24]. Ein verpflichtender Leistungsnachweis in ärztlicher Gesprächsführung ist in der Studienordnung einzelner Fakultäten festgeschrieben [25], gemäß der Approbationsordnung für Ärzte jedoch keine bundeseinheitliche Vorgabe [23]. Dass Anamnestizieren und Diagnostizieren dennoch als zentraler Bestandteil ärztlicher Kompetenz und als Grundvoraussetzung für die ärztliche Approbation in Deutschland angesehen werden, zeigt sich darin, dass Kompetenzen diesbezüglich im dritten Abschnitt der ärztlichen Prüfung explizit erfasst werden [23].

Die effektivste Möglichkeit Diagnosekompetenzen zu fördern ist Trainieren anhand realitätsgetreuer Problemstellungen [17]. Lernen anhand Patienten ist jedoch aus Sicherheits- und Kostengründen sowie aufgrund ethischer Faktoren eingeschränkt. In einer Studie des Instituts für Didaktik und Ausbildungsforschung in der Medizin am Klinikum der Ludwig-Maximilians-Universität München (DAM) gaben 25 % der befragten Studierenden an, bereits einen medizinischen Fehler während des Studiums begangen zu haben und 64 % wünschten sich in der Ausbildung mehr instruktionale Unterstützung [26]. Meta-Analysen zeigten, dass Simulationen eine

ideale Möglichkeit sind, Kompetenzen in der medizinischen Ausbildung zu trainieren und Diagnosekompetenzen hierbei nicht nur eingeübt, sondern auch effektiv vermittelt werden können [27].

Die bestehende Problematik der Fehldiagnosen in der Medizin, der zentrale Beitrag des Anamnesegegesprächs zur Diagnosefindung sowie die Notwendigkeit eines kontinuierlichen, aktiven Trainierens der Anamnesefertigkeiten rechtfertigen die kontinuierliche Evaluation von Messinstrumenten und simulierten Lernumgebungen. Ziel ist eine Verbesserung der Diagnostizierfähigkeit bereits im Medizinstudium, um damit den durch Fehldiagnosen begründeten Schaden an Patienten und Gesundheitssystem zu minimieren.

Theoretisches Faktenwissen von Medizinstudierenden kann anhand konzeptueller Wissenstests überprüft werden [28]. Aktives Trainieren mit klinischen Fallvignetten, beispielsweise durch Key Feature Tests oder Simulationen, dient dazu theoretisches Wissen praxisbezogen zu verknüpfen oder zu messen. Da Diagnostizieren ein komplexer Prozess und abhängig von Kontext und Lernumgebung ist, wird angenommen, dass Diagnosekompetenzen nicht durch ein einzelnes Messinstrument beurteilt werden können [29]. Im Rahmen der vorliegenden Doktorarbeit werden deshalb verschiedene Messinstrumente zur Ermittlung klinischer Diagnosekompetenzen in simulierten Lernumgebungen untersucht.

## 1.2 Theoretische Grundlagen

### 1.2.1 Begriffsklärungen

*Diagnostizieren* ist „zielgerichtetes Sammeln und Integrieren von Informationen [zum Zweck der Unsicherheitsreduktion]“ [30], fallspezifisch zur Beurteilung einer Situation oder für weiteres strategisches Vorgehen [31]. In der Medizin ist Diagnostizieren ein komplexer Prozess. Entscheidungen werden mehr individuell als standardisiert in einem multizentrierten, dynamischen Kontext getroffen [31], von einem Arzt alleine, aber auch kollaborativ im ärztlichen Team oder in Zusammenarbeit mit Patienten [31]. Symptome und Befunde werden einer Erkrankung zugeordnet.

Diagnostizieren erfolgt nicht nur in der Medizin, sondern beispielsweise auch in der Pädagogik. Beiden Fachbereichen ist gemeinsam, dass theoretisches Grundwissen, wissenschaftliche Erkenntnisse und individuelle Erfahrungen des Diagnostizierenden in die spezifische Situation integriert werden [32]. Pädagogische Diagnosen dienen jedoch nicht der Kategorisierung von

Befunden und Phänomenen wie in der Medizin, sondern der Optimierung von individuellen Lernprozessen. Relevante Lernvoraussetzungen werden zur gezielten Förderung ermittelt [33].

*Kompetenz* beschreibt individuelle Voraussetzungen einer Person zur erfolgreichen Bewältigung situationsspezifischer Anforderungen [34]. Kompetenz ist erlern- und vermittelbar [35] und kann hinsichtlich kognitiver, sowie handlungsbezogener Aspekte definiert werden. Kognitive Kompetenz ist die kontextabhängige Leistungsdisposition, erworben durch Lernen und dafür notwendig, Anforderungen in spezifischen Domänen zu bewältigen [34]. Handlungskompetenz beinhaltet zusätzlich metakognitive, motivationale und selbstregulative Merkmale einer Person, die als entscheidende Voraussetzungen für die Bereitschaft zu handeln gesehen werden [34]. Performanz beschreibt sichtbares und beobachtbares Verhalten, in dem sich Kompetenz zeigt und kann beispielsweise anhand von Testergebnissen in Messinstrumenten erfasst werden [36].

*Diagnosekompetenzen*, kognitive, kontextspezifische Leistungsdispositionen [37], zeigen sich auf Grundlage diagnostisch relevanten Wissens und diagnostischer Fertigkeiten in der Qualität von Diagnosen. Diese wird durch die Expertise des Diagnostizierenden, seinem Wissen und Fertigkeiten, beeinflusst. Expertise entwickelt sich, wenn Individuen Wissen, Verständnis und Können in ihrem Handlungsfeld dazugewinnen, dies kritisch bewerten und damit gezielt agieren [21]. Experten erbringen in ihrer Domäne dauerhaft und nicht nur einmalig oder zufällig herausragende Leistungen [19]. Hiervon abzugrenzen sind Neulinge oder Anfänger, Novizen. Der Unterschied zwischen beiden Gruppen zeigt sich nicht nur im Umfang von Wissen, Verständnis und Können, sondern vor allem in der Fähigkeit, diese Komponenten zu interferieren und anzuwenden. Ärzte mit langjähriger Berufserfahrung, wie Fachärzte, sind als Experten, Medizinstudierende als Novizen zu erachten [38]. Erste funktionale klinische Schemata der Wissensverknüpfung werden in der Intermediärstufe gebildet und erweitert [39] (s.1.3.1). Dies geschieht unter anderem während des Praktischen Jahres und im Rahmen der Facharztausbildung.

*Clinical Reasoning*, das klinische Denken, definiert kognitive Aspekte, und somit Denk- und Entscheidungsprozesse im Rahmen klinischer Tätigkeiten [40]. Neben den kognitiven sind fachliche und personenbezogene Faktoren bei klinischem Denken involviert. Kognitive Resultate zeigen sich als bestmöglichstes Vorgehen in einer bestimmten Situation, fachliche Aspekte beinhalten auch ethische und rechtliche Faktoren. Das personenbezogene Handeln zeigt sich in der Miteinbeziehung und respektvollen Zusammenarbeit mit Kollegen, Pflegekräften und Pati-

enten [40]. Clinical Reasoning bezieht sich auf den kompletten diagnostischen und therapeutischen Entscheidungsprozess, umfasst Anamnese, körperliche Untersuchungen und korrekte Interpretation der Befunde, um eine richtige Diagnose und Therapie effizient zu erreichen [41], [42]. Trowbridge et al. (2015) beschreiben Clinical Reasoning zusammenfassend als einen Vorgang, in dem Kliniker mit Patient und Umfeld interagieren, Patienteninformationen erhalten und Nutzen und Risiko von Handlungen abwägen, um eine Arbeitsdiagnose und einen Therapieplan zum Wohl des Patienten festzulegen [43]. Clinical Reasoning ist ein zyklischer, dynamischer Prozess, bei dem Anamnestizieren, Beobachtung und praktische Untersuchungsfertigkeiten mit epistemisch-diagnostischen Aktivitäten (s. 1.3.5) wie Hypothesengenerierung, Hypothesentestung, Analysieren, Synthetisieren miteinander verknüpft werden [44].

Je nach inhaltlichem Schwerpunkt werden in der Literatur verschiedene Bezeichnungen über Clinical Reasoning hinaus für klinisches Denken verwendet. Hierzu zählen Clinical Cognition, Medical Problem Solving, Medical Decision Making und Diagnostic Reasoning.

### 1.2.2 Professionelle Wissensbasis

Wissen basiert auf fachspezifischen Informationen und Daten, ist aber im Gegenteil zu diesen immer an eine Person gebunden [45]. Es wird dafür eingesetzt Probleme zu bearbeiten und zu lösen. Medizinisches Wissen wird in der Literatur uneinheitlich klassifiziert.

Mayer et al. (2010) unterscheiden 5 Arten von Wissen: Faktenwissen (Wissen bzgl. Tatsachen), konzeptuelles Wissen (Wissen bzgl. Ursachen und Wirkungen), prozedurales Wissen (Wissen bzgl. situationsspezifischem Vorgehen), strategisches Wissen (Wissen bzgl. allgemeinem Vorgehen) sowie die Einstellungshaltung (Selbstkonzept des Lernenden) [46].

Im Zusammenhang mit Diagnostizieren sind, in Anlehnung an Paris und Wixon (1983), konzeptuelles, strategisches und konditionales Wissen relevant [47], [48], [28]: Konzeptuelles Wissen wird auch als deklaratives Wissen, *knowing that*, bezeichnet. Dieses wissenschaftlich fundierte Faktenwissen von Definitionen und Diagnosekriterien dient dazu, Handlungsziele festzulegen und zu adaptieren [47]. Strategisches Wissen, *knowing how*, beinhaltet Informationen über bestmögliche Ausübungsstrategien für eine Handlung [47] sowie weiteres Vorgehen und dient somit der Festlegung diagnostischer Schritte. Konzeptuelles und strategisches Wissen sind gemäß der Forschergruppe des DAM die professionelle Wissensbasis medizinischer Diagnosekompetenzen [30] (s. 1.3.5). Konditionales Wissen, *knowing when* und *knowing why*, umfasst die richtigen Handlungsschritte unter gewissen Umständen [47].

Strategisches Wissen und konditionales Wissen beziehen sich auf Handlungsvorgehensweisen und werden als prozedurales Wissen zusammengefasst [28]. In Bezug auf Diagnosekompetenzen dient prozedurales Wissen dazu, schrittweise Probleme und medizinische Fragestellungen zu lösen.

Durch Reflexion und praktische Anwendung werden prozedurales und konzeptuelles Wissen in klinisches Wissen umgewandelt [49]. Klinisches Wissen ist somit eine Form der Wissensorganisation und besteht aus abgespeicherten Mustern und Regeln [50]. Es gibt Aufschluss darüber, wie sich Krankheiten bei einem Patienten manifestieren können [51]. Hiervon abzugrenzen ist biomedizinisches Wissen, das Wissen über die anatomischen, biochemischen und pathophysiologischen Prozesse, die einer Krankheit zu Grunde liegen [51].

## **1.3 Diagnosekompetenzen**

### **1.3.1 Entwicklung von Diagnosekompetenzen**

Diagnosekompetenzen entstehen mit dem Erwerb medizinischer Expertise. Patel und Groen (1991) beschreiben diesbezüglich 4 Entwicklungsstadien und unterscheiden zwischen Novizen, Intermediates, generischen Experten und Experten [52]. Gemäß der Stadientheorie medizinischer Expertise entstehen Diagnosekompetenzen gemeinsam mit dem Erwerb und Ausbau medizinischer Fertigkeiten und sind fest an theoretisches und praktisches Wissen gebunden [17].

Folgende Charakteristika kennzeichnen die 4 Entwicklungsstadien medizinischer Expertise: Zu Beginn der medizinischen Ausbildung verfügen Studierende über eine begrenzte Wissensbasis aus vornehmlich biomedizinischem Wissen und wenig klinischem Wissen [39]. Sie werden als Novizen (1) bezeichnet. Novizen erkennen lediglich isoliert Zeichen und Symptome klinischer Fallvignetten, die sie aus der Grundlage ihres biomedizinischen Wissens ableiten, und auf einzelne pathophysiologische Konzepte beziehen [38]. Informationen werden detailliert und bewusst verarbeitet, jedoch keine klinischen Muster erkannt. Das wiederholte Anwenden von erlerntem Wissen anhand von Patientenfällen führt zur Enkapsulierung des Wissens. Einzelne detaillierte Wissens Elemente biomedizinischen Wissens werden in größere Entitäten zusammengefasst und mit klinischem Wissen verknüpft. Wissen über pathophysiologische Prozesse wird dahingehend umgewandelt und erweitert, klinische Merkmale und Symptome zu erkennen [38]. Enkapsulierte Wissensnetzwerke sind Kennzeichen des Intermediates Stadiums (2), das

mit fortschreitender Studiumsdauer erreicht wird. Mit weiter zunehmenden klinischen Erfahrungen und Patientenkontakt werden im Stadium des generischen Experten (3) aus Wissensnetzwerken Illness Scripts gebildet.

Illness Scripts sind kognitive Einheiten aus klinisch relevanten Informationen für mögliche Manifestationsformen einer Erkrankung [38]. Sie bestehen aus 3 Komponenten: Enabling Conditions, den Umständen und Einflussfaktoren, Fault, den pathophysiologischen Prozessen, sowie Consequences, Konsequenzen pathophysiologischer Prozesse und somit den Symptomen einer Erkrankung [17].

Im Gegensatz zu Wissensnetzwerken bezieht sich ein Illness Script auf ein komplettes Krankheitsbild und wird nicht schrittweise, sondern im Ganzen aktiviert [17]. Wissen hinsichtlich Symptome und klinischer Parameter eines Krankheitsbildes sind dabei nicht exklusiv einem bestimmten Illness Script zugeordnet, sondern können in mehreren Scripts gleichzeitig vertreten sein [17]. Inhalt, Aufbau und Umfang eines Illness Scripts werden kontinuierlich durch eigenes Fachwissen und Erfahrungen ergänzt und überarbeitet. Der Einfluss dieser individuellen Faktoren erklärt individuell unterschiedliche Herangehensweisen bei der Bearbeitung klinischer Fragestellungen [53]. Zu Beginn eines Diagnoseprozesses sind oftmals mehrere Illness Scripts gleichzeitig aktiviert. Sie sind Basis dafür, welche weiteren Informationen durch Anamnese oder Untersuchungen einzuholen sind. Unter Einbeziehung dieser Ergebnisse werden manche Illness Scripts infolge wieder deaktiviert, andere bleiben weiter aktiv. Illness Scripts sind eine Kombination aus abgespeicherten Mustern und Patientenerfahrungen und Ausdruck der Wissensorganisation auf der Diagnosekompetenzen gründen [54].

Im Stadium des Experten (4) sind Illness Scripts um Patientenerfahrungen erweitert und werden als Instance Scripts abgespeichert [19]. Die Aktivierung von Instance Scripts eines Experten ist im Vergleich zur Aktivierung von Wissensnetzwerken eines Novizen ein automatischer, unbewusster Prozess [17]. Das biomedizinische Wissen als Wissensbasis ist bei Experten indes in einem Art Stand-By Modus und kann bei komplexen Fragestellungen eingesetzt werden [55].

### **1.3.2 Anwendung von Diagnosekompetenzen**

Diagnosekompetenzen werden im Kontext des Problemlösens medizinischer Fragestellungen und somit des klinischen Diagnostizierens angewendet. Über die inhaltliche Fachkompetenz hinaus fordern sie weitere Dimensionen der beruflichen Handlungskompetenz. Hierzu zählen kognitive Fähigkeiten (Methodenkompetenz), Sozialkompetenz und Personalkompetenz [56].



Strategisches Vorgehen im Rahmen des Diagnostizierens erfolgt entweder hypothetisch-deduktiv, somit analytisch, oder nicht-analytisch via Mustererkennung [5].

Das hypothetisch-deduktive Modell basiert auf der Bildung von Hypothesen, die aus der Kombination von spezifischen Merkmalen des Patienten mit gesammelten Daten aufgestellt werden [5]. Sie bilden die Grundlage für weiteres Vorgehen; Entscheidungen werden von den Ausgangshypothesen abgeleitet und sind somit deduktiv. Hypothetisch-deduktives Vorgehen beinhaltet 6 Schritte. Pre-assessment Image (1) beschreibt erste Assoziationen, die auf Basis von Vorinformationen bereits vor dem Patientenkontakt gebildet werden. Im Zuge der Cue Acquisition (2) durch das Anamnesegespräch werden Schlüsselinformationen erfasst, und daraus Hypothesen, häufig mehrere, gebildet, Hypothesis Generation (3). Bezogen auf den Prozess des Diagnostizierens sind dies letztlich Verdachtsdiagnosen. Durch Cue Interpretation (4) werden weitere Informationen hinsichtlich dieser Verdachtsdiagnosen gezielt gesammelt und den Hypothesen zugeordnet. Diese werden bewertet, Hypothesis Evaluation (5), und aus der am besten gesicherten Hypothese die Diagnose, Diagnosis (6), gebildet [50]. Hypothetisch-deduktives Vorgehen, als langsamer, analytischer Prozess ist dem kognitiven System 2 zuzuordnen [57] (s. 1.3.3).

Weisen Patientenfälle ausreichend saliente Merkmale auf, um darin klinische Muster, Pattern, zu erkennen, erfolgt Diagnostizieren durch nicht-analytisches Vorgehen [58]. Illness Scripts werden aktiviert und dienen dem weiteren Problemlöseprozess. Mustererkennung wird ab dem Stadium des generischen Experten angewendet, die auf Wissen und Vorerfahrungen aus vergleichbaren Situationen zurückgreifen [50]. Dieser Prozess ist durch Schnelligkeit, Effizienz und Intuition charakterisiert [57] und wird daher dem kognitiven System 1 zugeordnet (s. 1.3.3). Im Vergleich zum hypothetisch-deduktiven Vorgehen birgt dieser jedoch ein größeres Risiko für Fehler, wenn adäquates Wissen und Erfahrung fehlen [59].

In der Praxis ist eine klare Trennung beider strategischer Vorgehensweisen jedoch selten; Elemente der Mustererkennung werden auch bei Experten häufig mit analytischem Vorgehen kombiniert [57]. Fehlen Wissen oder Erfahrung, um klinische Muster zu erkennen, wie bei Novizen der Fall, erfolgt Diagnostizieren oft hypothetisch-deduktiv. Bei komplexen oder unbekannten Fallszenarien greifen auch Experten auf diese Form des Diagnostizierens zurück [5]. Durning et al. (2015) zeigten allerdings in einer unter Fachärzten für Innere Medizin durchgeführten Studie, dass durch nicht-analytisches Vorgehen auch bei der Beantwortung von als komplex empfundenen Fragen mehr korrekte Antworten erreicht wurden [57].

Kennzeichen verschiedener Expertisestadien in Bezug auf den Clinical Reasoning Prozess sind in Tabelle 1 zusammengefasst.

Tabelle 1

*Kennzeichen verschiedener Expertisestadien im Clinical Reasoning Prozess*

<b>Expertisestadium</b>	<b>Wissensorganisation</b>	<b>Strategisches Vorgehen</b>	<b>Kognitive Anforderung</b>
<b>Novize</b>	Wissensnetzwerke	Bildung von Hypothesen	Hoch
<b>Intermediate</b>	Enkapsuliertes Wissen	Anwendung von Wissensnetzwerken	Mittel
<b>Generische Experte</b>	Illness Scripts	Mustererkennung	Niedrig
<b>Experte</b>	Instance Scripts	Mustererkennung	Niedrig

### 1.3.3 Einflussfaktoren auf Diagnosekompetenzen

Diagnosekompetenzen werden von personenspezifischen sowie externen Faktoren beeinflusst. Personenspezifische Einflussfaktoren sind Wissen, Kognition und Metakognition.

Wissen stellt die inhaltliche Grundlage von Diagnosekompetenzen dar. Die verschiedenen Arten von Wissen in Bezug auf Diagnosekompetenzen, konzeptuelles Wissen, strategisches Wissen, und konditionales Wissen, sowie biomedizinisches und klinisches Wissen wurden im Abschnitt 1.2.2 dargestellt. Der Wissensumfang erhöht die Wahrscheinlichkeit, eine korrekte Entscheidung zu treffen, ist alleine jedoch kein aussagekräftiger Prädiktor für Diagnosekompetenzen [50]. Wissen steht in wechselseitiger Beziehung zu Kognition und wirkt sich, kritisch angewendet, positiv auf Diagnosekompetenzen aus [18]. Kritisches Denken, Critical Thinking, bedeutet, Informationen, Situationen, Ursachen und Perspektiven zu analysieren [60] und erfordert Reflexion und offene Auseinandersetzung mit eigenem Wissen [18].

Es gilt, kognitive Fähigkeiten von kognitiven Fertigkeiten zu unterscheiden. In Bezug auf Diagnosekompetenzen sind beispielsweise epistemisch-diagnostizierende Aktivitäten wie Hypothesenbildung, Evaluation von Evidenzen und Schlussfolgerungen ziehen (s. 1.3.5) als kognitive Fähigkeiten zu erachten. Aus der Vernetzung dieser Fähigkeiten können dann kognitive Fertigkeiten, wie Strategien zur Lösung komplexer Problemstellungen, entstehen. Kognitive Fähigkeiten bilden somit die Basis für kognitive Fertigkeiten [61]. Gemäß Kahneman (2011) wird Kognition über zwei verschiedene Systeme angewendet, die als System 1 und System 2 bezeichnet werden [62]. Denken im Rahmen des System 1 ist ein schneller, automatischer, instinktiver, unbewusster Vorgang, der geringe kognitive Anstrengung erfordert. Der Denkprozess im System 2 ist langsam, bewusst, regel-basiert und kontrolliert und bedingt somit kognitiver

Anstrengung [62]. Welches der beiden Systeme im Rahmen des Diagnostizierens angewendet wird, ist von Merkmalen eines Patientenfalls aber auch der klinischen Erfahrung des Diagnostikers abhängig. Oftmals werden beide Systeme gleichzeitig aktiviert und System 1 und System 2 deshalb als duales Prozessmodell bezeichnet [5].

Metakognition ist die „Wahrnehmung(en) der eigenen Überlegungen, also die Fähigkeit sich das eigene Denken bewusst zu machen oder darüber nachzudenken“ [63]. Als eine schwer messbare Fähigkeit ist sie dem Wissen und der Kognition überzuordnen. Eine Form von Metakognition ist Reflection in Action (deutsch: Reflexion während einer Handlung) [64]. Sie ist „Grundlage dafür, dass der Lernende allgemeines strategisches Wissen aus spezifischen Lernaktivitäten herleitet“ [65] und kann als Bindeglied zwischen Denken und Handeln erachtet werden. Durch die Reflexion können klinische Ergebnisse besser verstanden und optimiert werden [63]. Metakognitive Fähigkeiten sind elementarer Einflussfaktor für Diagnosekompetenzen und als *das* Kennzeichen medizinischer Expertise zu werten [44]. Novizen reflektieren seltener und hauptsächlich über das Verhalten von Patienten, wohingegen Experten häufiger und nicht nur Patienten sondern auch eigenes Verhalten beurteilen [66].

Von extern wirken Situative Faktoren, Scaffolding und Case Specificity auf Diagnosekompetenzen ein.

Ein Situativer Faktor ist die Komplexität eines Patientenfalls. Diese ergibt sich aus der Anzahl möglicher Differentialdiagnosen, Begleiterkrankungen oder der Länge der Anamnese eines Patienten. Des Weiteren beeinflussen Umfeld und Umstände der Diagnosesituation Diagnosekompetenzen: Zeitdruck, Stress, die Menge an verfügbaren Patientendaten sowie Möglichkeiten weitere Untersuchungen anzuordnen fließen in den Diagnoseprozess mit ein. Durch Scaffolding wird die Aufmerksamkeit auf wichtige fallspezifische Aspekte gelenkt mit dem Ziel, Lernenden das Problemlösen zu erleichtern [67]. Instruktionale Unterstützungen ermöglichen den Umfang an Cognitive Load auf Lernende zu modulieren. Einer Überforderung des Arbeitsgedächtnis und somit negativen Auswirkungen auf den weiteren Lernerfolg kann hierdurch vorgebeugt [68] und der Erwerb von Diagnosekompetenzen gefördert werden [27]. Scaffolding beeinflusst den Erwerb, Case Specificity die Messung von Diagnosekompetenzen. Gemäß Annahmen zur Case Specificity sind Diagnosekompetenzen nicht generalisierbar. Diagnosestrategien sind situationsspezifisch und die Diagnose eine individuelle Entscheidung für jeden einzelnen Patientenfall [21].

Die Auswirkungen von personenspezifischen und externen Einflussfaktoren zeigen sich bei Patienten anhand des Behandlungserfolgs und der Zufriedenheit. Beim Diagnostizierenden selbst

zeigen sich potenzielle Effekte im Arbeitsablauf, der interkollegialen Kommunikation aber auch der eigenen Zufriedenheit. Im Gesundheitssystem zeigen sich Diagnosekompetenzen in ökonomischen Kennzahlen und den daraus resultierenden Konsequenzen (s. Abbildung 1).

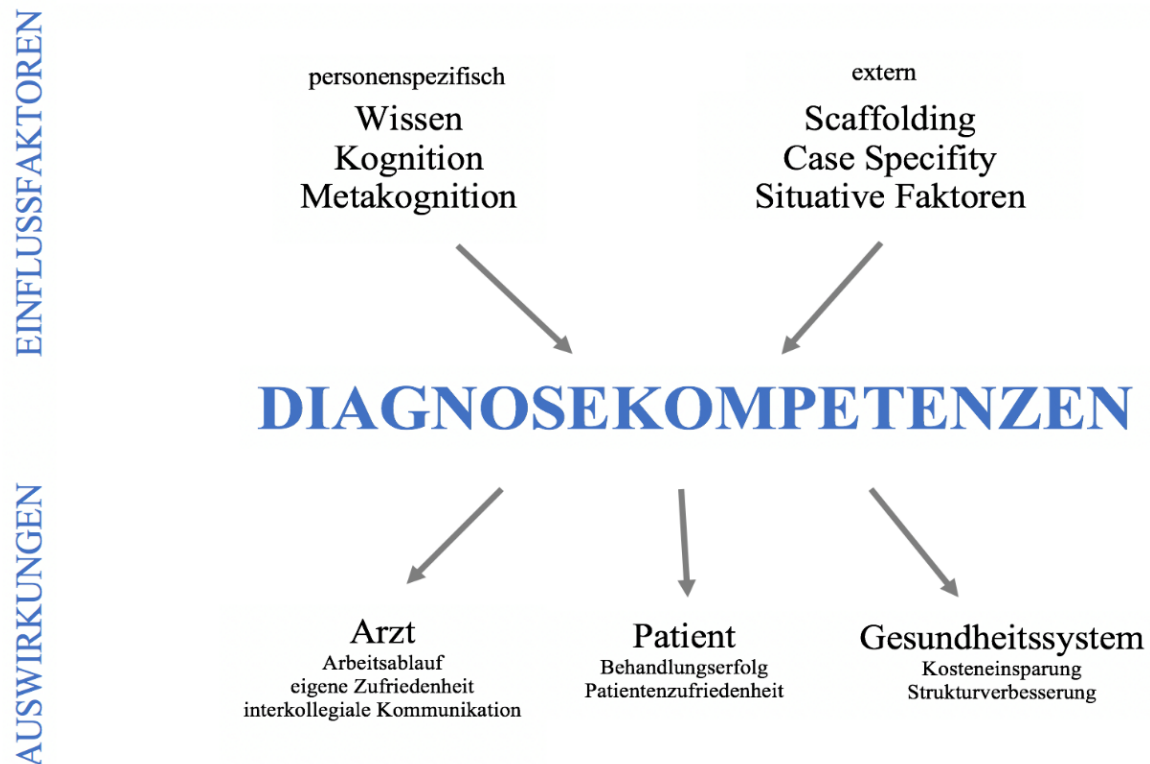


Abbildung 1. Einflussfaktoren und Auswirkungen von Diagnosekompetenzen  
Quelle: Eigene Darstellung

### 1.3.4 Lernumgebungen für Diagnosekompetenzen

Die Anwendung von Wissen in realitätsgetreuen Lernumgebungen wird als essentiell für die Entwicklung und Förderung von Diagnosekompetenzen erachtet [69]. Simulationen sind eine authentische Gelegenheit für den Erwerb klinischer Qualifikationen [70] und ermöglichen es, Wissen sowie praktische und klinische Fertigkeiten zu zeigen und anzuwenden [22]. Die Co-sima Forschergruppe der LMU definiert Simulationen als Lernumgebungen, in denen „(1) Realitätsausschnitte angemessen mit Blick auf den Zweck der Diagnose dargestellt werden und (2) die Lernenden in die Situation eingreifen können [... durch tatsächliches Handeln oder das Treffen von Entscheidungen]“ [30]. Durch Simulationen können Problemstellungen authentisch vereinfacht dargestellt werden und Situationen, die sehr selten, zu gefährlich oder zu teuer sind, um erstmalig in Realität ausgeübt zu werden, trainiert werden [71].

Lernen im Rahmen von Simulationen bietet stark kontextbezogene praktische Erfahrungen in experimentellen Umgebungen [22]. In der medizinischen Ausbildung wird simulationsbasiertes Lernen häufig in Skills Labs angewendet [72]. Der Fokus hierbei liegt auf dem Erwerb praktischer Fertigkeiten. Jedoch können auch theoretische Fähigkeiten wie Diagnosekompetenzen effektiv in Simulationen gefördert werden [8]. Anamnestizieren, als Ausgangsschritt im Diagnoseprozess, kann nicht nur im Rahmen traditioneller Lehrveranstaltungen wie Seminare, sondern auch anhand von Simulationen gelehrt werden [8]. Diese ermöglichen den Umfang an Cognitive Load gezielt zu regulieren [22] und somit den Lernerfolg für den Erwerb von Diagnosekompetenzen positiv zu beeinflussen. Durch Fiction Contracts werden Teilnehmende mit den Bedingungen einer Simulation vertraut gemacht und in die Lernumgebung eingeführt [73].

Live-Simulationen werden beispielsweise mit standardisierten Patienten durchgeführt. Standardisierte Patienten stellen einen Patienten derart präzise und authentisch dar, dass dies in der Situation von einem erfahrenen Kliniker nicht als Simulation zu erkennen ist [22]. Zusätzlich zu den körperlichen Symptomen werden standardisierte Patienten darauf vorbereitet, Emotionen auszudrücken, Stress zu vermitteln und Erwartungen an den Arzt zu richten. Sie ermöglichen unmittelbares Feedback und Reflexion [22]. Die Durchführung von Live-Simulationen ist höchst interaktiv, jedoch an die Verfügbarkeit von Räumlichkeiten und Schauspielern gebunden sowie kostenintensiv. Darüber hinaus bedingen sie einen hohen administrativen Aufwand.

Aus Effizienzgründen muss simulationsbasiertes Lernen jedoch kostengünstiger, zugänglicher und besser replizierbar sein [22]. Einen Lösungsansatz diesbezüglich bieten computerbasierte Video-Simulationen, welche bereits Ende der 1990er Jahre im medizinischen Curriculum eingesetzt wurden [74]. Anderweitig schwer replizierbare klinische Szenarien können standardisiert dargestellt werden und darüber hinaus Lernfortschritt sowie häufige Fehler ermittelt werden [8]. Der operationelle Unterschied zu Live-Simulationen besteht in der Technologie des Lerndesigns: Die Kommunikation erfolgt über Menü- oder Freitexteingaben, die im Anschluss analysiert werden [75]. Der inhaltliche Unterschied zeigt sich bei den ebenso standardisierten Patienten, die jedoch lediglich Symptome aufweisen ohne situationsspezifisch Reaktionen zeigen zu können. Video-Simulationen dienen der standardisierten Replizierbarkeit von klinischen Szenarien [8]. Die Konzeption von Video-Simulationen ist an hohe technische Anforderungen gebunden, jedoch können diese unbegrenzt oft sowie zeit- und ortsunabhängig in Online-Lernumgebungen verwendet werden.

### 1.3.5 Arbeitsmodell für Diagnosekompetenzen

Die theoretische Grundlage der vorliegenden Arbeit ist das Arbeitsmodell für Diagnosekompetenzen. Es wurde von der DFG-Forscherguppe Cosima der LMU zur „Förderung von Diagnosekompetenzen in simulationsbasierten Lernumgebungen in der Hochschule“ entwickelt [30] (s. Abbildung 2).

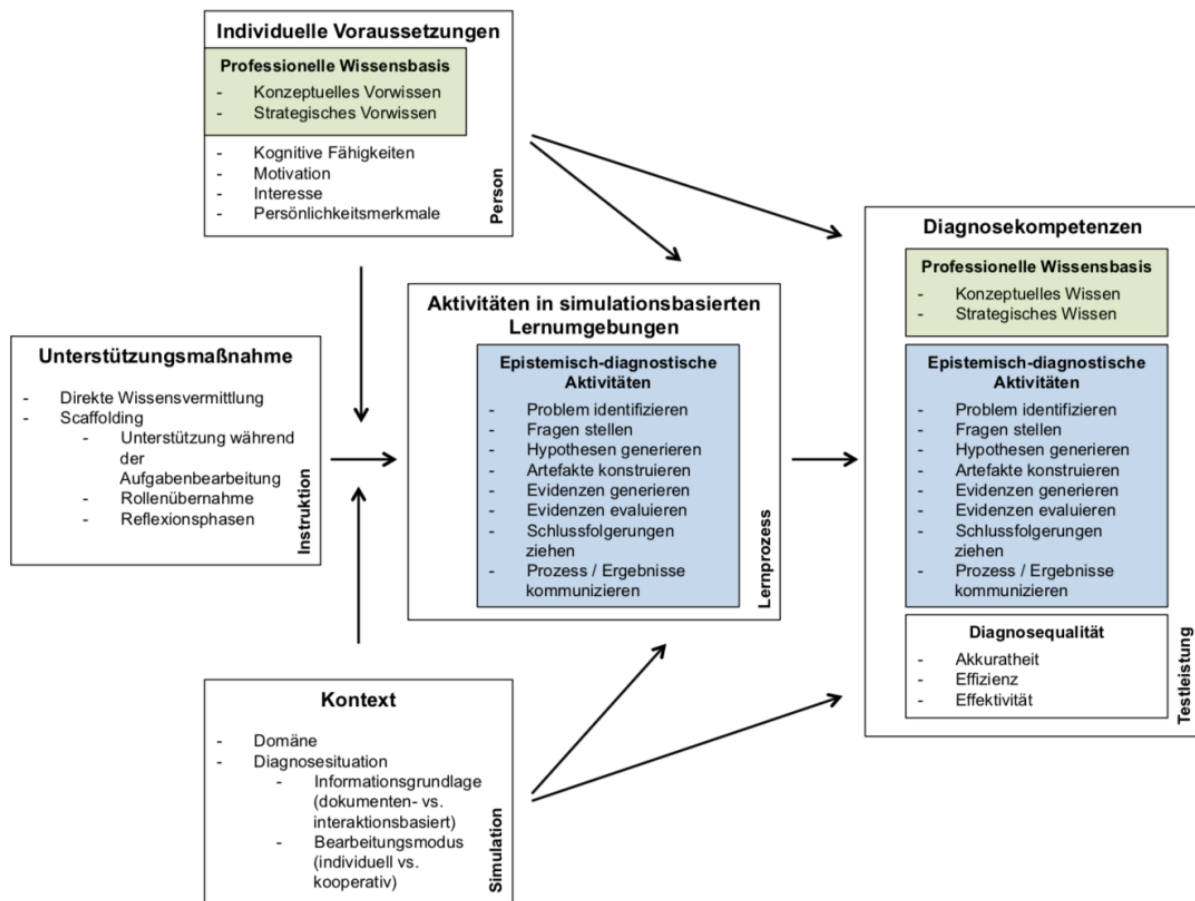


Abbildung 2. Arbeitsmodell für Diagnosekompetenzen der DFG-Forscherguppe COSIMA  
Quelle: Heitzmann, N. et al. (angenommen, Überarbeitungen ausstehend)

Diesem zufolge setzen sich Diagnosekompetenzen aus der professionellen Wissensbasis, epistemisch-diagnostischen Aktivitäten und Diagnosequalität zusammen.

Die professionelle Wissensbasis, strategisches und konzeptuelles Wissen (s.1.2.2), ist Teil der individuellen Voraussetzungen für Diagnosekompetenzen. Diese beeinflussen darüber hinaus den Lernprozess sowie das Ausmaß, in dem sich Unterstützungsmaßnahmen positiv auf den Lernprozess auswirken.

Zweites Element von Diagnosekompetenzen sind 8 epistemisch-diagnostische Aktivitäten, welche in simulationsbasierten Lernumgebungen ausgeführt und erfasst werden können. Diese

sind: Problem identifizieren, Fragen stellen, Hypothesen generieren, Artefakte konstruieren, Evidenzen generieren, Evidenzen evaluieren, Schlussfolgerungen ziehen sowie Prozess/Ergebnisse kommunizieren.

Im Rahmen des Diagnostizierens bedeutet Problem identifizieren ein Symptom zu erkennen, Fragen stellen die Ursachen eines Symptoms zu hinterfragen und Hypothesen generieren eine Verdachtsdiagnose aufzustellen. Auf Basis der Verdachtsdiagnose werden weitere Untersuchungen angeordnet, womit Evidenzen generiert werden. Evidenzen zu evaluieren bedeutet Untersuchungsergebnisse auszuwerten, woraus Schlussfolgerungen wie das weitere therapeutische Vorgehen getroffen werden. Abschließend werden diese Maßnahmen mitgeteilt und erklärt. Die Ausführung und die Qualität der einzelnen Aktivitäten sind von den individuellen Voraussetzungen einer Person, Unterstützungsmaßnahmen sowie kontextspezifischen Faktoren abhängig.

Diagnosekompetenzen zeigen sich in der Diagnosequalität. Diese setzt sich dem Arbeitsmodell zufolge aus den Performanzmaßen Akkuratheit, Effizienz und Effektivität zusammen. Sie können beispielsweise anhand von Testleistungen in Simulationen ermittelt werden.

### **1.3.6 Messinstrumente für Diagnosekompetenzen**

Diagnosekompetenzen, ein multidimensionales Konstrukt aus professioneller Wissensbasis, epistemisch-diagnostischen Aktivitäten sowie Diagnosequalität [30], können nicht durch ein einzelnes Messinstrument ausreichend quantifiziert werden [29]. Die Messung erfolgt anhand von Wissen und Verhalten in einer spezifischen Situation in einem mehrstufigen, kontinuierlichen Prozess [76]. Kompetenzen in einzelnen Teilbereichen werden durch verschiedene spezifische Testinstrumente bewertet. Wissenstests dienen der Evaluation der professionellen Wissensbasis. Durch praktische Tests können sowohl Wissen als auch Fertigkeiten überprüft werden.

Gemäß der Pyramide von Miller werden in der medizinischen Ausbildung Diagnosekompetenzen anhand der Fertigkeiten *knows*, *knows how*, *shows how* und *does* gemessen [77] und können mittels spezifischer Messinstrumente erfasst werden (s. Abbildung 3). Deskriptives Fachwissen (*knows*) bildet die Basis der Pyramide und kann durch Multiple Choice (MC)-Fragen, Handlungswissen (*knows how*) durch Key Feature Tests geprüft werden. Verhalten wird in simulierten Situationen (*shows how*) mittels OSCE (Objective Structured Clinical Examination) und in realen Situationen (*does*) bewertet [78].

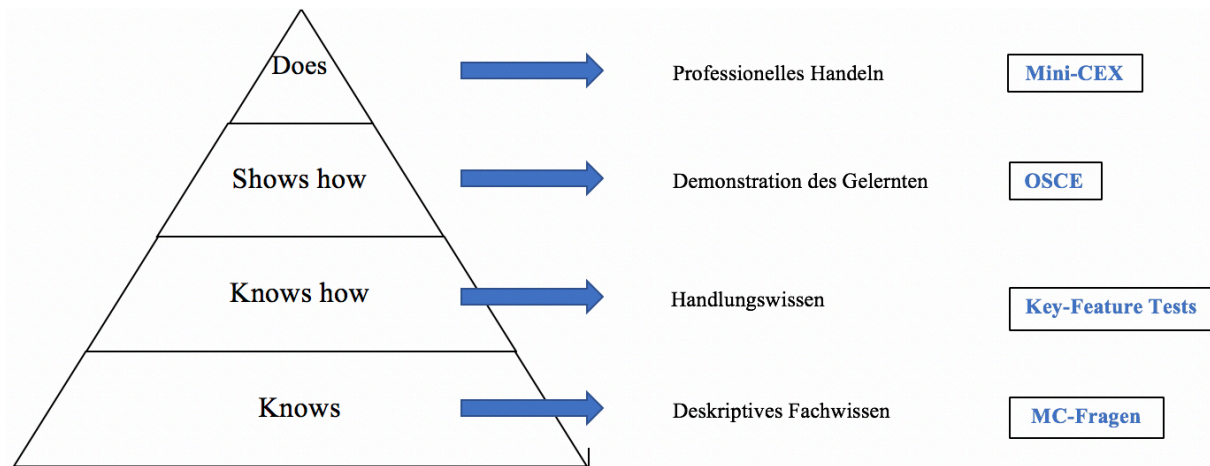


Abbildung 3. Modifizierte Wissenspyramide nach Miller und Messinstrumente zur Erfassung einzelner Teilbereiche von Diagnosekompetenzen  
Quelle: Eigene Darstellung

Key Feature Tests sind ein etabliertes Instrument, um strategisches Wissen zu prüfen [28]. Sie liefern aussagekräftige Ergebnisse hinsichtlich der Qualität ärztlicher Leistung und Patiententoutcome und können somit für die ökonomische Messung von Diagnosekompetenzen effektiv eingesetzt werden [79]. Der Ausdruck Key Feature (KF) wurde von Page et al. (1995) als *kritischer Schritt* in Zusammenhang mit der Erfassung von Entscheidungskompetenzen etabliert [80]. Ein Key Feature ist zum einen der entscheidende Aspekt, der zur Identifikation und Lösung von Problemen beiträgt, und fokussiert zum anderen den Schritt, bei dem Lernende am häufigsten Fehler in der Lösung von Problemen begehen [78]. Die Grundidee von Key Features ist, die Aufmerksamkeit auf die essentiellen Aspekte klinischer Fälle zu richten, um dadurch effizienter Wissen prüfen zu können [81]. Sie bestehen aus der „knappen Darstellung einer klinischen Situation, dem sog. Stamm, gefolgt von drei bis fünf Fragen, den KF-Fragen“ [78].

Mittels Konzeptueller Wissenstests wird deklaratives Wissen hinsichtlich problembasierter Fragestellungen gemessen und zusätzlich kritisches Denken, Wissenstransfer und Eigenreflexion während der Fallbearbeitung verbessert [82]. Die Testkonzeption ist meist ein Multiple Choice (MC) Format, wobei eine oder mehrere Antworten aus einer vorgegebenen Anzahl von Antwortmöglichkeiten auszuwählen sind [83]. Darüber hinaus dienen Flashcards, Lernkarten, der Überprüfung von konzeptuellem Wissen [28].

Key Feature Tests und Konzeptuelle Wissenstests messen die professionelle Wissensbasis von Diagnosekompetenzen. In stattgehabten Untersuchungen korrelierten Testergebnisse der beiden Wissenstests lediglich schwach positiv [28]. Dies wird dadurch erklärt, dass bei fallbasier-



ten Problemstellungen, wie in Key Feature Tests, andere kognitive Prozesse aktiviert werden als bei fakten- und problemorientierten Fragestellungen [84].

Simulationen dienen nicht nur als Lernumgebung für Diagnosekompetenzen, sondern auch als Messinstrument. Beispielsweise im Rahmen eines OSCE werden in simulationsbasierter Lernumgebung anhand standardisierter Patienten verschiedene Dimensionen klinischer Kompetenzen integriert erfasst [85]. In Abgrenzung zu schriftlichen Prüfungsformaten werden hierbei neben Wissen und Denken auch klinische Kompetenzen, wie eben das Führen von Anamnesegesprächen oder praktische Fertigkeiten, wie körperliche Untersuchungen, geprüft [86]. Live-Simulationen mit standardisierten Patienten als Messinstrument für klinische Kompetenzen zeigen eine höhere Validität und Reliabilität als mündlichen Prüfungen bei einzelnen Patienten [87]. Die Validität und Reliabilität von Video-Simulationen hängt von instruktionalen Unterstützungsmaßnahmen, die den Cognitive Load und die Leistung beeinflussen können, ab [88]. Obgleich durch das OSCE elementare Fertigkeiten von Diagnosekompetenzen gemessen werden, wurde gezeigt, dass diese dadurch nicht ausreichend ganzheitlich erfasst werden [85], und somit hierfür mehrere Messinstrumente heranzuziehen sind.

## 1.4 Ziel der Arbeit und Fragestellungen

Ziel der vorliegenden Arbeit ist, Instrumente zur Messung von klinischen Diagnosekompetenzen bei Medizinstudierenden zu validieren. Dies dient der Optimierung von Lern- und Testumgebungen für das interaktionsbasierte Diagnostizieren im Kontext des medizinischen Anamnesegesprächs.

Als Messinstrumente wurden 2 Wissenstests, Konzeptueller Wissenstest und Key Feature Test, sowie Live- und Video-Simulationen für Anamnesegespräche untersucht. Sie wurden für das Teilprojekt 4 „Förderung von interaktiven Diagnosekompetenzen im simulierten Anamnesegespräch: Effekte der übernommenen Rolle und unterschiedlicher Reflexionsphasen“ der DFG-Forschergruppe Cosima der LMU konzipiert.

Das Forschungsprojekt besteht aus 4 Studienphasen (s. Anhang 1). Die in der vorliegenden Arbeit ausgewerteten Daten wurden in der Studienphase 1 zur Entwicklung und Optimierung des Studienmaterials erhoben.

Über die Validierung der Messinstrumente hinaus werden folgende Fragestellungen in dieser Arbeit untersucht:

1. Wie korreliert die Performanz im Konzeptuellen Wissenstest mit der Performanz im Key Feature Test?
2. Wie korreliert die Performanz in den Wissenstests mit der Performanz in den Simulationen?
3. Welche Faktoren sind mit der Performanz in Simulationen assoziiert?
  - a. Art der Simulation (Live-Simulation, Video-Simulation)
  - b. Effekt der Reihenfolge (Live- vor Video-Simulation bzw. Video- vor Live-Simulation)?

## 2 Methodik

### 2.1 Studiendesign

Es wurde eine Studie mit dem Titel „Üben von Anamnesegesprächen in Simulationen“ am Zentrum für Unterricht und Studium (ZeUS) der LMU im Wintersemester 2018/2019 durchgeführt. Die Erhebungen fanden an 12 Terminen unter der Leitung von Maximilian Fink, Pädagoge, M.Sc. der Psychologie und Doktorand am DAM, statt.

Inhaltsdomäne war das Leitsymptom Atemnot. Atemnot als Ausdruck vorwiegend kardiopulmonaler Erkrankungen ist aufgrund der Häufigkeit des Auftretens ein hoch relevanter Konsultationsanlass [89], der einen umfassenden Fokus in der Ausbildung von Medizinstudierenden rechtfertigt. Angesichts der mannigfaltigen Ätiologien und Manifestationsformen von Atemnot ist hierfür eine detaillierte und gezielte Anamnese notwendig, um effiziente und effektive diagnostische Entscheidungen treffen zu können.

Das Studiendesign bestand aus einer Prätest-Phase und einer Simulations-Phase. In der Prätest-Phase wurden die Wissenstests, Konzeptueller Wissenstest und Key Feature Test durchgeführt. In der Simulationsphase wurde im Anschluss an das Anamnesegespräch eine virtuelle Patientenakte bearbeitet. Die Simulationsphase war ein 2 x 2-Design mit den Variablen Simulationsart (Live- und Video-Simulation) als within-subject Faktor und Reihenfolge als between-subject Faktor (s. 2.4). Zusätzlich waren in beiden Phasen pädagogisch-psychologische Fragebögen implementiert. Die Gesamtdauer der Studie betrug 4,5 Stunden (s. Tabelle 2).

Tabelle 2  
*Ablauf der Studie*

Ablauf Briefing		Prätestphase		Simulationsphase		Debriefing
		Konzeptueller Wissenstest	Key Feature Test	Live-Simulation	Video-Simulation	
Zeit	10 Min.	40 Min.	40 Min.	85 Min.	85 Min.	10 Min.

## 2.2 Studienteilnehmende

Teilnehmende waren Medizinstudierende der LMU und TU München. Die Rekrutierung erfolgte über Aushänge an den Universitätsgebäuden, Online über die Homepages der medizinischen Fakultäten sowie soziale Netzwerke. Teilnahmevoraussetzung war der Studiumsstatus: Medizinstudierende im Klinischen Studienabschnitt (S) oder Praktischen Jahr (P). Die Teilnahme an der Studie war freiwillig, wurde mit 45€ vergütet und mit einer schriftlichen Einverständniserklärung zum Datenschutz bestätigt. Für die Studie wurde von der Ethikkommission der LMU eine Unbedenklichkeitsbescheinigung ausgestellt.

Insgesamt nahmen 92 Medizinstudierende an der Studie teil. 86 Testergebnisse gingen in die Auswertung ein, 6 wurden aufgrund unvollständiger Erhebungen oder Teilnahme an der Pilotstudie nachträglich ausgeschlossen. 47 waren Studierende im Klinischen Studienabschnitt (S), wovon 40 bereits eine Famulatur absolvierten, 39 waren Studierende im Praktischen Jahr (P). Die Teilnehmenden wurden gebeten, ihr Wissen zu Atemnot und Anamnese anhand einer 5-stufigen Likert Skala einzuschätzen. Für Atemnot ergab sich als Mittelwert 3.12, für Anamnese 3.47 Punkte. Studierende im Praktischen Jahr schätzten ihr Wissen im Mittel höher ein als Studierende im Klinischen Studienabschnitt. Tabelle 3 zeigt die demographischen Angaben des ausgewerteten Kollektivs.

Tabelle 3

*Demographische Angaben des ausgewerteten Studienkollektivs*

		absolut	relativ (%)	M	SD	Min	Max
<b>Geschlecht</b>	m	32	37				
	w	54	63				
<b>Alter</b>				26.03	4.72	21	56
<b>Ausbildung im Gesundheitsbereich</b>		16	19				
<b>Studiumsstatus</b>	(S) Klinischer Studienabschnitt	47	55				
	(P) Praktisches Jahr	39	45				
<b>Ausbildungsstatus (S)</b>	kardiovaskuläres System	35	75				
	respiratorisches System	34	72				
	Notfallmedizin	32	68				

		absolut	relativ (%)	M	SD	Min	Max
<b>Anamnesegruppe</b>	Visitenkompetenz	15	32				
	Famulatur	40	85				
		9	11				
		(S) 6 (P) 3	(S) 13 (P) 8				
<b>Wissens- einschätzung</b>	<b>Atemnot</b>			3.12 (S) 2.77 (P) 3.55	0.80 (S) 0.70 (P) 0.72	1 (S) 1 (P) 2	5 (S) 4 (P) 5
	1-gering	2 (S) 2 (P) 0	2 (S) 4 (P) 0				
	etwas	14 (S) 12 (P) 2	16 (S) 26 (P) 5				
	mittel	45 (S) 28 (P) 17	52 (S) 60 (P) 44				
	viel	22 (S) 5 (P) 17	26 (S) 11 (P) 44				
	5-sehr viel	3 (S) 0 (P) 3	4 (S) 0 (P) 8				
	<b>Anamnese</b>			3.47 (S) 3.23 (P) 3.76	0.68 (S) 0.56 (P) 0.71	2 (S) 2 (P) 2	5 (S) 4 (P) 5
	1-gering	0 (S) 0 (P) 0	0 (S) 0 (P) 0				
	etwas	4 (S) 3 (P) 1	5 (S) 6 (P) 3				
	mittel	43 (S) 30 (P) 13	50 (S) 64 (P) 33				
	viel	34 (S) 14 (P) 20	40 (S) 30 (P) 51				
	5-sehr viel	5 (S) 0 (P) 5	6 (S) 0 (P) 13				

## 2.3 Studienmaterial

Das Studienmaterial wurde in CASUS, einer Online-Plattform für Lehr- und Trainingszwecke in der studentischen Ausbildung, implementiert. Der Zugriff auf CASUS erfolgte über den Standard Webbrowser Mozilla Firefox.

Zur Validierung des Studienmaterials wurde 2018 eine Pilotstudie mit 12 Medizinstudierenden im Klinischen Studienabschnitt durchgeführt. Hierbei wurden Konzeptueller Wissenstest, Simulationen und pädagogisch-psychologische Messungen überprüft. Der Key Feature Test wurde nachträglich konzipiert und den Leitern der Arbeitsgruppen des DAM zur inhaltlichen Revision zur Verfügung gestellt.

Qualitätsindices für das Studienmaterial waren die Gütekriterien Objektivität, Reliabilität und Validität [90]. Objektivität, die Unabhängigkeit der Testergebnisse vom Prüfer, wurde im Rahmen der Messung durch die computerbasierte Implementierung der Wissenstests sowie standardisierte Bedingungen in den Simulationen erreicht. Ex ante festgelegte Musterlösungen gewährleisteten die Objektivität der Auswertung. Reliabilität, die Reproduzierbarkeit der Testergebnisse, wurde anhand der internen Konsistenz mit Cronbachs  $\alpha$  bestimmt. Die Konzeption der Wissenstests im Multiple Choice Antwortformat verspricht eine hohe Reliabilität [86]. Konzeptueller Wissenstest und Key Feature Test bestanden jeweils aus 40 Fragen. Diese Anzahl ist zur Gewährleistung der Reliabilität von Messungen im Multiple Choice Antwortformat notwendig [91]. Standardisierte Fragestellungen garantierten Reliabilität der Testergebnisse aus den Simulationen. Für die Validität, die Messspezifität, wurden aus dem Studienmaterial Item-Schwierigkeit, interne Konsistenz in den Wissenstests sowie die Korrelationen der Testergebnisse verschiedener Messinstrumente bestimmt. Darüber hinaus wurde Validität durch den Einsatz der Messinstrumente in realitätsgetreuem Umfeld mit Simulationen erreicht [92].

### **2.3.1 Inhalt des Studienmaterials**

#### **2.3.1.1 Konzeptueller Wissenstest**

Der Konzeptuelle Wissenstest bestand aus 40 Fragen (=Frageitem). Diese wurden von Doktoranden und Mitarbeitern des DAM zu den Krankheitsbildern Lungenembolie, Herzinsuffizienz, Pneumonie, COPD, sowie psychosomatische und sonstige Ursachen von Atemnot erstellt. 26 Fragen waren im Single Choice Typ A Format („Bitte kreuzen Sie eine Antwort an“), 14 Fragen im Pick-N Multiple Choice Format („Bitte kreuzen Sie jede zutreffende Aussage an“) konzipiert. Es standen jeweils 5 Antwortmöglichkeiten (=Antwortitem) zur Auswahl.

Exemplarisch sind Fragen aus dem Konzeptuellen Wissenstest als Anhang 2 beigelegt.

#### **2.3.1.2 Key Feature Test**

Der Key Feature Test bestand aus 10 Fällen (=Fallitem) mit jeweils 4 Fragen (=Frageitem), je einer zu Verdachtsdiagnose, Symptom, Untersuchung und Therapie. Die Fragen wurden von Doktoranden und Mitarbeitern des DAM zu den Krankheitsbildern Blutungsanämie, Glottisödem, COPD, Pneumonie, Lungenembolie, Pneumothorax, Herzinsuffizienz und Asthma erstellt. Alle Fragen waren im Single Choice Typ A Format („Bitte kreuzen Sie eine Antwort an“) konzipiert. Es standen jeweils 12 Antwortmöglichkeiten (=Antwortitem) zur Auswahl.

Exemplarisch ist ein Key Feature Fall als Anhang 3 beigelegt.

#### **2.3.1.3 Simulationen**

Die Simulationen bestanden aus 6 Patientenfällen zu den Krankheitsbildern Lungenembolie (2), Herzinsuffizienz (2), Pneumonie und Hyperventilationstetanie. Im Anschluss an das Anamnesegepräch wurde in einer virtuellen Patientenakte Verdachtsdiagnose, Begründung der Verdachtsdiagnose, Differentialdiagnosen, Untersuchungen und Therapiemaßnahmen angegeben. Verdachtsdiagnose und Differentialdiagnosen wurden aus einer Liste über ein Drop-Down Menü mit 239 Items ausgewählt. Begründung der Verdachtsdiagnose sowie Maßnahmen für Untersuchungen und Therapie erfolgten in Freitexteingabe.

Die Fälle wurden von Frau Dr. med. Hannah Gerstenkorn, Fachärztin für Allgemeinmedizin und wissenschaftliche Mitarbeiterin am DAM erstellt. 3 Fälle wurden von weiblichen Schauspielern, 3 Fälle von männlichen Schauspielern dargestellt. Jedem Schauspieler wurde ex ante ein Patientenfall zugeteilt, den er sowohl bei der Aufzeichnung der Video-Simulation als auch in Live darstellte.

Die Patientenfälle werden im weiteren Verlauf der Arbeit mit den Nachnamen der simulierten Patienten bezeichnet (s. Tabelle 4).

Tabelle 4

*Diagnosen der Patientenfälle aus den Simulationen*

Patientenname	Geschlecht	Diagnose
Aimüller	w	Lungenembolie bei Gerinnungsstörung
Albrecht	m	Herzinsuffizienz bei Hypertropher obstruktiver Kardiomyopathie (HOCM)
Klein	w	Pneumonie
Wagner	m	Lungenembolie bei Prostatakarzinom
Bircher	w	Herzinsuffizienz bei Tachyarrhythmia absoluta
Lehner	m	Hyperventilation bei Panikattacke

Alle Fälle waren gleich strukturiert, für jeden Fall wurde eine Fallvignette konzipiert. Diese bestand aus einem Einführungstext mit Hintergrundinformationen für die Schauspieler, einem Patientenprotokoll mit Vorbefunden, als Basisinformation für die Medizinstudierenden, und 69 standardisierten, fallidentischen Anamnesefragen. Die Anamnesefragen waren in 5 Kategorien aufgeteilt: Hauptsymptome, Medizinische Vorgeschichte, Allergien und Aktuelle Medikation, Sozial- und Familienanamnese sowie Systemübersicht. Im Voraus wurde von der Autorin für jede standardisierte Frage eine fallspezifische Musterantwort festgelegt. Diese Frage-Antwort-Paare waren Drehbuchgrundlage für die Aufzeichnung der Video-Simulation sowie Orientierungsgrundlage für die Schauspieler zur Beantwortung der Anamnesefragen in der Live-Simulation. Für die Video-Simulation wurde jede Frage als Einzelsequenz unter ärztlicher Supervision in Laborräumen des DAM aufgezeichnet. Die Live-Simulationen fanden im Rahmen der Studie im ZeUS statt. Die standardisierte Einrichtung der Räumlichkeiten gewährleistete eine Vergleichbarkeit beider Settings. Für die Live-Simulation erhielten die Schauspieler die Anweisung, sich bei der Beantwortung der Anamnesefragen an die vorhandenen Antworten aus der Fallvignette zu orientieren oder bei Unklarheiten „Das weiß ich nicht“ zu antworten. Eine ärztliche Supervision der Schauspieler während der Live-Simulation fand nicht statt.

Exemplarisch ist die klinische Fallvignette eines Patientenfalls als Anhang 4 beigelegt.



#### **2.3.1.4 Pädagogisch-psychologische Messungen**

Die Pädagogisch-psychologischen Messungen bestanden aus Fragebögen zu Selbstwirksamkeit, Cognitive Load, Sicherheit, Spatial Presence, Emotionen, Interesse, Motivation und Authentizität der Lernumgebung. Die Ergebnisse dieser Messungen sind nicht Bestandteil der vorliegenden Arbeit und werden im Rahmen einer psychologischen Doktorarbeit verwendet.

### **2.3.2 Kodierung der Messinstrumente**

#### **2.3.2.1 Konzeptueller Wissenstest**

Der Konzeptuelle Wissenstest wurde in Anlehnung an Bauer et al. (2011) ausgewertet [93]. Single Choice Typ A Fragen wurden dichotom bewertet: Jede gemäß der Musterlösung korrekt beantwortete Frage wurde mit 1 Punkt bewertet, falsche Antworten mit 0 Punkten. Pick-N Multiple Choice Fragen wurden partiell nach der PS<sub>50</sub> Methode ausgewertet: 100 % korrekt beantwortete Items wurden mit 1 Punkt, 50 % oder mehr korrekt beantwortete mit 0.5 Punkte, weniger als 50 % korrekt beantwortet Items mit 0 Punkten bewertet. Nicht beantwortete Fragen wurden mit 0 Punkten bewertet. Das Testergebnis lag somit zwischen 0 und 40 Punkten.

#### **2.3.2.2 Key Feature Test**

Die Auswertung des Key Feature Tests erfolgte gemäß Bauer et al. (2011) dichotom mittels Ganzpunktbewertung [93]: Jede gemäß der Musterlösung korrekt beantwortete Einzelfrage wurde mit 1 Punkt bewertet. Falsche Antworten und nicht beantwortete Fragen wurden mit 0 Punkten bewertet. Das Testergebnis lag somit zwischen 0 und 40 Punkten.

#### **2.3.2.3 Simulationen**

In der vorliegenden Arbeit wurden aus den Erhebungen der Simulationen Verdachtsdiagnose und Begründung der Verdachtsdiagnose ausgewertet. Für jeden Fall wurden im Voraus von der Fallautorin und einer weiteren Ärztin, unter Zuhilfenahme der Fallvignette, die korrekten Diagnosen sowie die relevanten Begründungsisems (klinische Befunde, Schlüsselbegriffe) unabhängig festgelegt. Übereinstimmende Diagnosen und Begründungsisems wurden als Musterlösung definiert (s. Anhang 5). Korrekte Diagnosen wurden mit ganzen Punkten bewertet (s. Tabelle 5). Begründungen wurden lediglich bei korrekten Diagnosen ausgewertet. Fallspezifisch lag die Anzahl relevanter Begründungsisems zwischen 5 und 10.

Tabelle 5

*Punktesystem zur Auswertung der Diagnosen der Patientenfälle aus den Simulationen*

<b>Punkte</b>	<b>Definition</b>
2	Eine der ex ante festgelegten Diagnosen wurde ausgewählt (d.h. eine Form des Krankheitsbildes, die der Fallkonzeption entspricht, wurde erkannt)
1	<ul style="list-style-type: none"> <li>■ Die ausgewählte Diagnose kann aufgrund der Informationen aus der Fallvignette erklärt werden</li> <li>■ Die ausgewählte Diagnose ist Teil der ex ante festgelegten Diagnosen, hätte aber aufgrund der verfügbaren Antwortmöglichkeiten aus der Liste genauer spezifiziert werden können.</li> </ul>
0	<ul style="list-style-type: none"> <li>■ Eine inkorrekte Diagnose wurde genannt</li> <li>■ Ein Symptom/ Untersuchungsbefund wurde genannt</li> <li>■ Das Diagnosefeld ist leer</li> </ul>

Die Auswertung der Begründung erfolgte gemäß folgendem Schema:

1. *Auswertung der Antworten:* Jedes aus der Musterlösung genannte Begründungsitem wurde mit 1 Punkt bewertet.
2. *Punktevergabe:* Der Quotient aus Anzahl genannter zu relevanten Begründungsitems ergab den jeweils fallspezifischen Punktescore.

Das Punkteintervall bei der Begründung der Verdachtsdiagnose lag dementsprechend zwischen 0 und 1 Punkt.

Pro Patientenfall konnten somit zwischen 0 und 3 Punkten erreicht werden. Das Testergebnis in den Simulationen lag zwischen 0 und 15 Punkte, da ein Patientenfall in der Auswertung ausgeschlossen wurde (s. 4.3).

## 2.4 Studiendurchführung

Zu Beginn der Studie wurde jedem Teilnehmenden eine PIN zugeteilt, die für die gesamte Erhebung der Identifikation in CASUS diene. Der erste Teil der Studie fand im Computerraum des ZeUS statt. In der Prätest-Phase wurden Konzeptueller Wissenstest, Key Feature Test sowie pädagogisch-psychologische Fragebögen bearbeitet. Im Anschluss folgte die Simulations-Phase mit der Erhebung der Anamnese und Bearbeitung der Patientenakte: Die Video-Simulationen fanden im Computerraum statt, die Live-Simulationen wurden in standardisiert, gleich

den Video-Aufzeichnungen, eingerichteten Räumen im ZeUS durchgeführt. Die Patientenfälle wurden gruppiert (Aimüller, Albrecht, Klein bzw. Wagner, Bircher, Lehner). Jeder Teilnehmende bearbeitete alle Patientenfälle, 3 in Live-Simulation, 3 in Video-Simulation. An einem Erhebungstermin wurde die Anamnese bei einer Gruppierung, entweder Gruppe 1 (Wagner, Bircher, Lehner) oder Gruppe 2 (Aimüller, Albrecht, Klein) in Live-Simulation erhoben, die der anderen Gruppierung in Video-Simulation. Die Zuteilung hinsichtlich der Startbedingung Live-Simulation (A) oder Video-Simulation (B) erfolgte per Losverfahren. Nach einer Familiarisierung hinsichtlich Simulationsbedingung war der Ablauf für jeden Patientenfall gleich. Die Teilnehmenden hatten über CASUS Zugriff auf ein virtuelles Patientenprotokoll, erhoben die Anamnese und bearbeiteten im Anschluss eine virtuelle Patientenakte.

Im Rahmen der Video-Simulation wurden die Anamnesefragen aus einem Auswahlménü gewählt. Dieses beinhaltete die 5 Fragekategorien Hauptsymptome, Medizinische Vorgeschichte, Allergien und aktuelle Medikation, Sozial- und Familienanamnese sowie Systemübersicht. Nach jeder Auswahl wurde die entsprechende Videosequenz mit der standardisierten Patientenantwort abgespielt. Innerhalb des Zeitfensters von 8:30 Min., welches mit dem Öffnen des Patientenprotokolls begann, konnten beliebig viele Anamnesefragen ausgewählt werden. Im Anschluss wurde das Auswahlménü beendet und die Patientenakte bearbeitet.

Im Rahmen der Live-Simulation wurde der jeweilige Patient in den Simulationsraum gebeten und nahm auf einer Liege Platz. Im Anschluss wurde das Patientenprotokoll geöffnet und die Zeitmessung gestartet. Das maximale Zeitfenster betrug 8:30 Min., das Anamnesegespräch konnte aber früher beendet werden. Die inhaltliche Gestaltung war frei. Nach Abschluss verließ der Patient den Simulationsraum und die Patientenakte wurde bearbeitet. Zur Orientierung und für Notizen stand den Studierenden ein Blanco-Anamnesebogen, welcher die 5 Fragekategorien gelistet hatte, zur Verfügung (s. Anhang 6).

Von den 86 ausgewerteten Testergebnissen wurden 40 in Gruppe 1, davon 19 in Bedingung A, 21 in Bedingung B erhoben. 46 Testergebnisse stammen aus Gruppe 2, davon 22 aus Bedingung A sowie 24 aus Bedingung B (s. Tabelle 6).

Tabelle 6

*Aufteilung der Testergebnisse aus den Simulationen nach Gruppe, Startbedingung und Studiumsstatus*

	Start Live-Simulation	Start Video-Simulation
<b>Gruppe 1</b>	1A	1B
<b>(Wagner, Bircher, Lehner)</b>	n = 19: (S) = 10, (P) = 9	n = 21: (S) = 12, (P) = 9
<b>Gruppe 2</b>	2A	2B
<b>(Aimüller, Albrecht, Klein)</b>	n = 22: (S) = 12, (P) = 10	n = 24: (S) = 13; (P) = 11

## 2.5 Statistische Analysen

### 2.5.1 Deskriptive Auswertung der Messinstrumente

Die deskriptive Auswertung beinhaltete eine Item- und Skalenanalyse der Wissenstests sowie die Berechnung der Testergebnisse der Messinstrumente unter Angabe von Häufigkeiten, Lage- und Streumaßen. Für die Berechnung der Testergebnisse und der Fragestellungen wurde der Datensatz zusätzlich anhand der kategorialen Variable Studiumsstatus in 2 Gruppen aufgeteilt: Studierende im Klinischen Studienabschnitt (S) sowie Studierende im Praktischen Jahr (P). Die Testergebnisse aus Konzeptuellem Wissenstest, Key Feature Test und Simulationen lagen als metrische Variablen vor.

Die Güte von Konzeptuellen Wissenstest und Key Feature Test wurde im Rahmen der Item- und Skalenanalyse untersucht. Jede Frage wurde als einzelnes Item betrachtet. Es wurden Schwierigkeit und Trennschärfe der Items berechnet, die Fragen wurden kategorisiert und die Reliabilität der einzelnen Subskalen ermittelt. Die Schwierigkeit einer Frage wurde über den Mittelwert der erreichten Punkte definiert [93]. Die Beurteilung von Trennschärfe und Reliabilität erfolgte anhand der Bildung von Skalen. Im Key Feature Test waren dies 4 Subskalen, Verdachtsdiagnose, Symptomatik, Diagnostik und Therapie, mit jeweils 10 Fragen. Im Konzeptuellen Wissenstest erfolgte die Zuordnung zu Symptomatik (8), Pathophysiologische Mechanismen (10), Diagnostik (10), Therapie (10) und Kategorie übergreifende Fragen (4). Die Trennschärfe wurde als Korrelation des Items mit dem zugehörigen Skalenwert exklusiv dem jeweiligen Item berechnet. Die Reliabilität wurde durch die interne Konsistenz mit Cronbachs  $\alpha$  bestimmt.

Die Auswertung der Messinstrumente erfolgte anhand der Berechnung von Mittelwerten sowie Lage- und Streumaße. Im Voraus wurde die Verteilung der Testergebnisse mittels Berechnung

von Schiefe und Kurtosis sowie graphisch anhand von Box-Plots untersucht. Die Messinstrumente wurden anhand von t-Tests hinsichtlich der kategorialen Variable Studiumsstatus überprüft.

### **2.5.2 Auswertung der Fragestellungen**

Zur Beantwortung der Fragestellungen 1 und 2 wurden bivariate Korrelationen unter Berechnung des Pearson Korrelationskoeffizienten durchgeführt und die Gruppierungen der Patientenfälle hinsichtlich Simulationsart berücksichtigt. Die Auswertung der dritten Fragestellung basierte auf einem logistischen Regressionsmodell für jeden Patientenfall, wobei die Entscheidung über die Diagnose als binäre Zielgröße aufgefasst wurde (korrekt/falsch). Neben den Haupteffekten Simulationsart und Startbedingung wurden auch jeweils die Interaktionseffekte in Betracht gezogen. Für den Effekt der Reihenfolge wurde untersucht, ob das Vorschalten von Video-Simulationen die Performanz in Live-Simulationen bzw. das Vorschalten von Live-Simulationen die Performanz in Video-Simulationen beeinflusst. Die Auswertungen erfolgten fallspezifisch aufgrund der diesbezüglichen Unterschiede in Inhalt und Schwierigkeit. Deskriptiv wurden hierfür gruppierte Ergebnisse aus Simulationsart und Startbedingung (Gruppen 1A, 1B, 2A, 2B) betrachtet.

Bei allen statistischen Analysen wurde das Signifikanzniveau mit 5 % festgesetzt.

Die Auswertung der Daten sowie die Grafikerstellung erfolgten mittels „SPSS-Statistics“ (Version 25.0) und „R“ (Version 3.5.0) in Zusammenarbeit mit dem StaBLab der LMU München.

### 3 Ergebnisse

#### 3.1 Deskriptive Auswertung der Messinstrumente

##### 3.1.1 Item-und Skalenanalyse der Wissenstests

###### Konzeptueller Wissenstest

Die Ergebnisse der Itemanalyse des Konzeptuellen Wissenstests sind in Tabelle 7 und Abbildung 4 dargestellt. 10 Fragen mit einer Trennschärfe  $< 0.1$  wurden für die weiteren Auswertungen der Testergebnisse ausgeschlossen. Die Reliabilität der einzelnen Skalen ( $\alpha_{\geq 0.1}$ ) wurde dadurch verbessert und die Trennschärfe der restlichen Items (3) verändert.

Tabelle 7

*Schwierigkeit und Trennschärfe der Frageitems im Konzeptuellen Wissenstest (1 = Schwierigkeit, 2 = Trennschärfe, 3 = Trennschärfe nach Ausschluss der Frageitems mit Trennschärfe  $< 0.1$ ; kursive Frageitems: Pick-N Multiple Choice Antwortformat)*

Symptomatik	KWQ1	KWQ6	KWQ16	KWQ27	KWQ29	KWQ37	KWQ38	KWQ40		
$\alpha = .30$										
$\alpha_{\geq 0.1} = .47$										
1	.88	.79	.48	.47	.64	.55	.31	.64		
2	.20	.16	.05	.24	.21	.27	-.08	.14		
3	.27	.25		.27	.20	.36		.12		
Pathophysiologische Mechanismen	KWQ3	KWQ5	KWQ7	KWQ11	KWQ18	KWQ20	KWQ24	KWQ26	KWQ31	KWQ33
$\alpha = .38$										
$\alpha_{\geq 0.1} = .46$										
1	.12	.28	.90	.30	.84	.53	.84	.90	.72	.26
2	.04	-.04	.26	.08	.19	.17	.23	.23	.28	.16
3			.24		.21	.17	.27	.27	.27	.16
Diagnostik	KWQ2	KWQ9	KWQ14	KWQ17	KWQ19	KWQ22	KWQ28	KWQ32	KWQ35	KWQ39
$\alpha = .42$										
$\alpha_{\geq 0.1} = .44$										
1	.91	.52	.95	.10	.84	.93	.77	.79	.15	.44
2	.12	.21	.25	.15	.14	.19	.15	.19	.03	.27
3	.11	.21	.24	.19	.17	.17	.18	.19		.24
Therapie	KWQ4	KWQ8	KWQ12	KWQ13	KWQ15	KWQ25	KWQ34	KWQ36		
$\alpha = .38$										
$\alpha_{\geq 0.1} = .47$										
1	.45	.55	.10	.43	.73	.40	.55	.65		
2	.09	.27	.06	.38	.26	.19	.12	-.04		
3		.34		.28	.31	.19	.17			

Kategorie	KWQ10	KWQ21	KWQ23	KWQ30
übergreifende Fragen				
$\alpha = .32$				
$\alpha_{\geq 0.1} = .41$				
1	.64	.78	.52	.66
2	.23	.03	.22	.20
3	.19		.22	.32

Die Reliabilität des Gesamttests betrug  $\alpha = .74$  und erreichte nach Ausschluss der Frageitems mit einer Trennschärfe  $< 0.1$   $\alpha_{\geq 0.1} = .76$ .

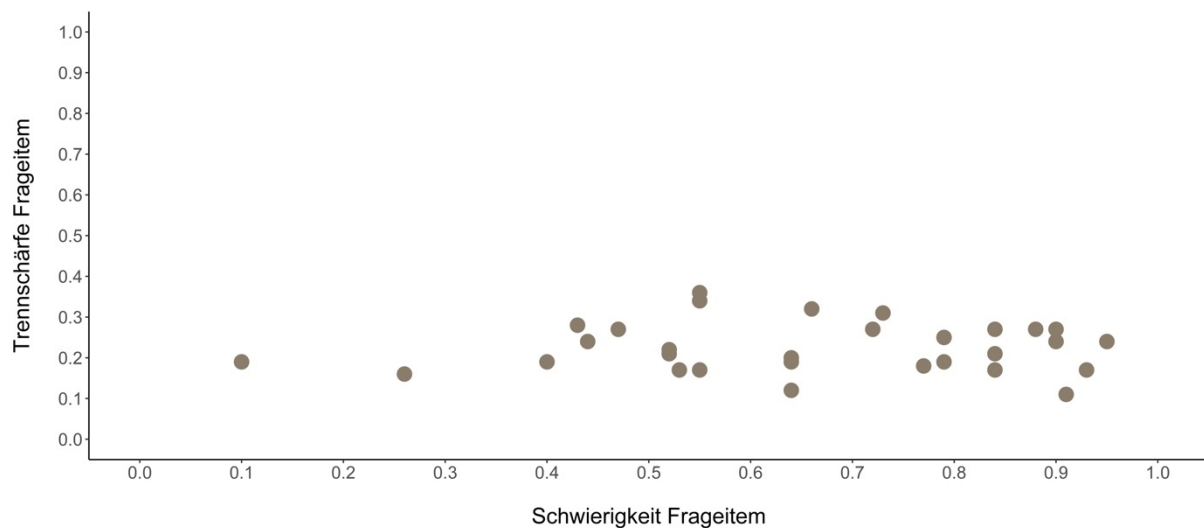


Abbildung 4. Schwierigkeit und Trennschärfe der ausgewerteten Frageitems ( $n = 30$ ) im Konzeptuellen Wissenstest

## Key Feature Test

Die Ergebnisse aus der Fallanalyse der Key Feature Fälle sind in Tabelle 8 und Abbildung 5 dargestellt.

Tabelle 8

*Schwierigkeit und Trennschärfe der Fallitems im Key Feature Test (1 = Schwierigkeit, 2 = Trennschärfe)*

KF	KF1	KF2	KF3	KF4	KF5	KF6	KF7	KF8	KF9	KF10
1	.56	.33	.43	.42	.65	.93	.88	.77	.65	.62
2	.40	.48	.20	.28	.37	.53	.51	.48	.55	.32

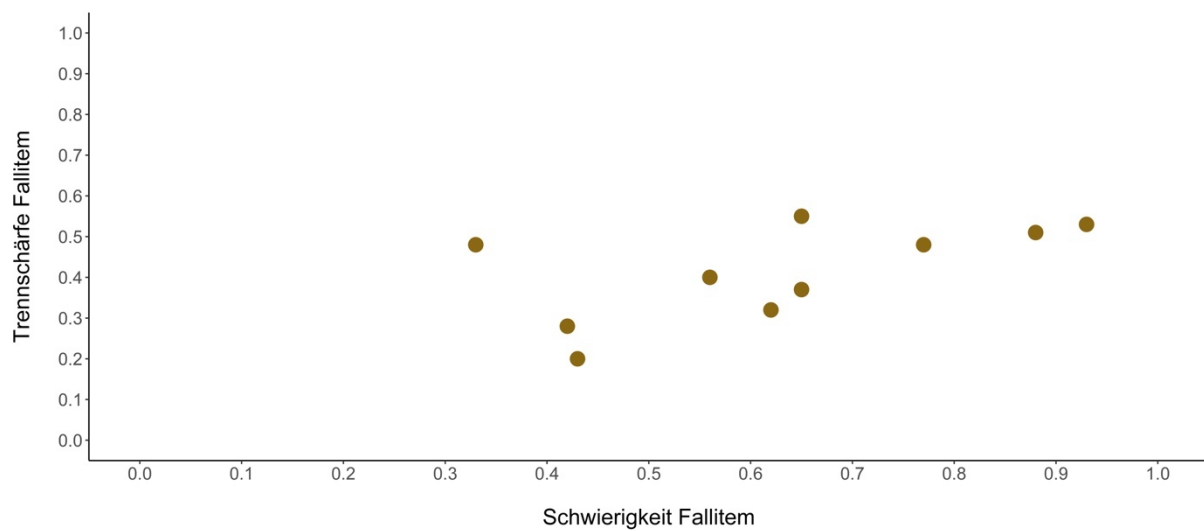


Abbildung 5. Schwierigkeit und Trennschärfe der Fallitems im Key Feature Test



Die Ergebnisse der Itemanalyse des Key Feature Tests sind in Tabelle 9 und Abbildung 6 dargestellt. 8 Fragen mit einer Trennschärfe  $< 0.1$  wurden für die weiteren Auswertungen der Testergebnisse ausgeschlossen. Die Reliabilität der einzelnen Skalen ( $\alpha_{\geq 0.1}$ ) wurde dadurch verbessert und die Trennschärfe der restlichen Items (3) verändert.

Tabelle 9

*Schwierigkeit und Trennschärfe der Frageitems im Key Feature Test (1 = Schwierigkeit, 2 = Trennschärfe, 3 = Trennschärfe nach Ausschluss der Frageitems mit Trennschärfe  $< 0.1$ )*

<b>Verdachts- diagnose</b>	<b>KFQ1</b>	<b>KFQ6</b>	<b>KFQ9</b>	<b>KFQ13</b>	<b>KFQ17</b>	<b>KFQ22</b>	<b>KFQ26</b>	<b>KFQ31</b>	<b>KFQ34</b>	<b>KFQ38</b>
$\alpha = .44$ $\alpha_{\geq 0.1} = .44$										
1	.74	.13	.79	.85	.94	.99	.95	.98	.85	.84
2	.16	.14	.28	.22	.06	.18	.15	.33	.25	.11
3	.12	.12	.32	.22		.19	.17	.34	.28	.08
<b>Symptomatik</b>	<b>KFQ2</b>	<b>KFQ5</b>	<b>KFQ10</b>	<b>KFQ14</b>	<b>KFQ18</b>	<b>KFQ21</b>	<b>KFQ25</b>	<b>KFQ29</b>	<b>KFQ33</b>	<b>KFQ37</b>
$\alpha = .49$ $\alpha_{\geq 0.1} = .55$										
1	.36	.38	.69	.17	.50	.76	.74	.85	.30	.23
2	.15	.14	-.07	.20	.34	.32	.39	.22	.30	.12
3	.14	.24		.23	.35	.30	.38	.20	.32	.12
<b>Untersu- chung</b>	<b>KFQ3</b>	<b>KFQ7</b>	<b>KFQ11</b>	<b>KFQ15</b>	<b>KFQ19</b>	<b>KFQ23</b>	<b>KFQ27</b>	<b>KFQ30</b>	<b>KFQ35</b>	<b>KFQ39</b>
$\alpha = .55$ $\alpha_{\geq 0.1} = .63$										
1	.70	.53	.14	.01	.69	.98	.84	.76	.64	.64
2	.23	.46	-.23	.07	.17	.35	.27	.52	.34	.28
3	.29	.50			.17	.33	.26	.53	.37	.30
<b>Therapie</b>	<b>KFQ4</b>	<b>KFQ8</b>	<b>KFQ12</b>	<b>KFQ16</b>	<b>KFQ20</b>	<b>KFQ24</b>	<b>KFQ28</b>	<b>KFQ32</b>	<b>KFQ36</b>	<b>KFQ40</b>
$\alpha = .26$ $\alpha_{\geq 0.1} = .38$										
1	.45	.27	.12	.65	.49	.99	.98	.50	.80	.76
2	.19	.22	-.1	-.08	.06	.30	.27	.11	.28	-.02
3	.26	.15				.30	.21	.15	.21	

Die Reliabilität des Gesamttests betrug  $\alpha = .79$  und erreichte nach Ausschluss der Frageitems mit einer Trennschärfe  $< 0.1$   $\alpha_{\geq 0.1} = .81$ .

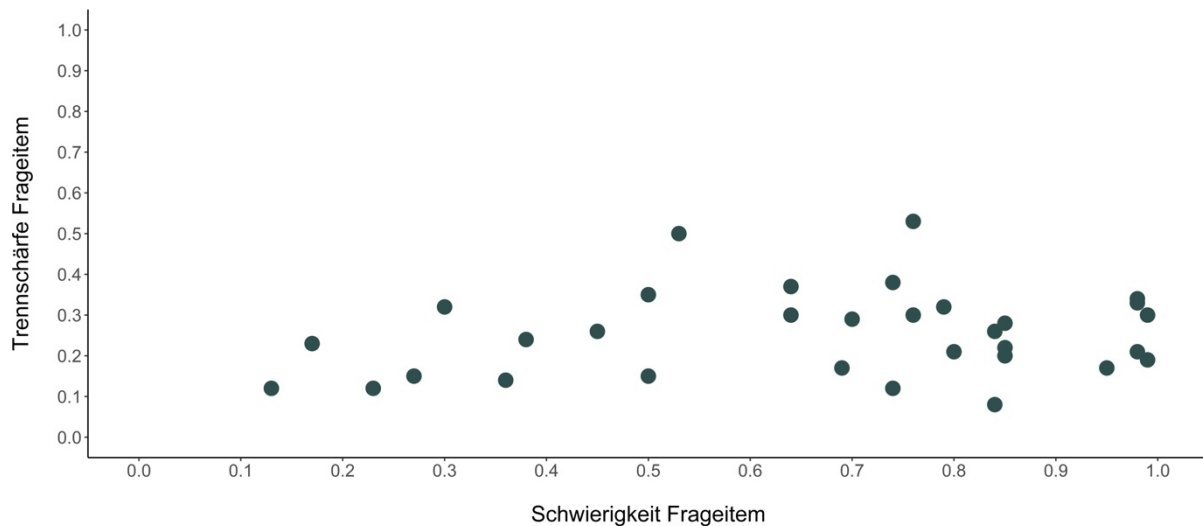


Abbildung 6. Schwierigkeit und Trennschärfe der ausgewerteten Frageitems ( $n = 32$ ) im Key Feature Test

### 3.1.2 Testergebnisse der Messinstrumente

Die Verteilung der Testergebnisse der Messinstrumente ist in Abbildung 7 dargestellt. Gemäß der in der Literatur angegebenen Richtlinien wurden keine Ausreißer identifiziert [94] und die Testergebnisse aller 86 ausgewerteten Teilnehmenden für die weiterführenden Berechnungen verwendet.

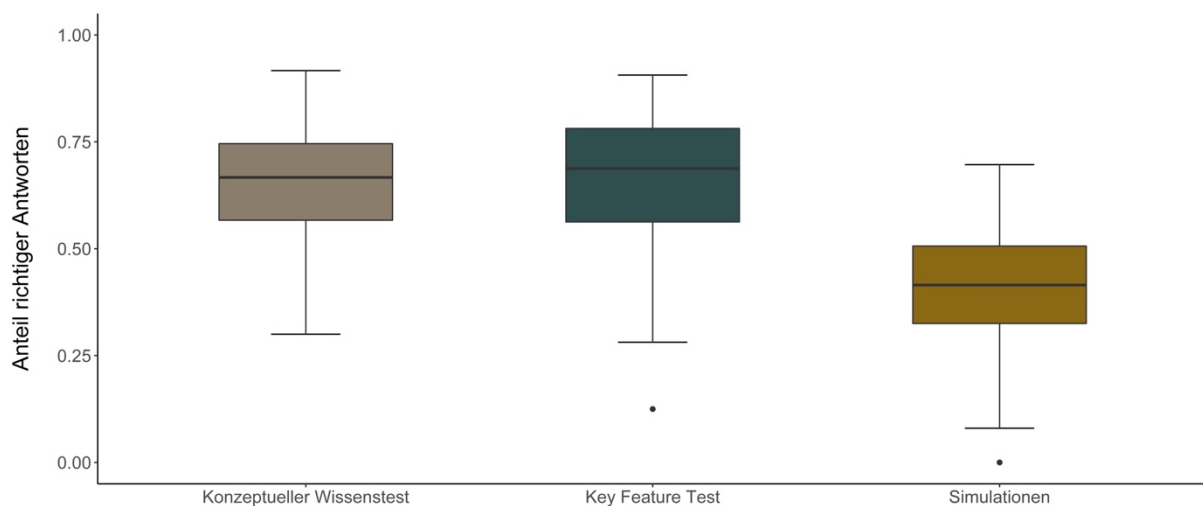


Abbildung 7. Verteilung der Testergebnisse der Messinstrumente

Die Überprüfung der Messinstrumente hinsichtlich einer Differenzierung des Studiumsstatus in den Testergebnissen ist im Anhang beigefügt (s. Anhang 7). Mittels 2-seitigem t-Test zwischen den ungepaarten Gruppen S und P ergab sich im Konzeptuellen Wissenstest ( $t(81) = 5.14, p < .01$ ) und im Key Feature Test ( $t(84) = 4.82, p < .01$ ) ein statistisch signifikanter

Unterschied in den Testergebnissen. Obgleich in den Simulationen ( $t(84) = 0.90, p = .37$ ) anhand der arithmetischen Mittel eine Tendenz zugunsten der Studierenden im Praktischen Jahr festzustellen ist (s. Tabelle 12), konnte auf dem Signifikanzniveau von 5 % keine allgemeingültige Differenzierung des Studiumsstatus in den Testergebnissen angenommen werden.

### Konzeptueller Wissenstest

Im Konzeptuellen Wissenstests war das durchschnittliche Testergebnis 66 %; 60 % bei Studierenden im Klinischen Studienabschnitt und 72 % bei Studierenden im Praktischen Jahr (s. Tabelle 10).

Tabelle 10

*Auswertung der Testergebnisse im Konzeptuellen Wissenstest*

<b>M</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>
.66	.13	.30	.92
(S) .60	(S) .11	(S) .30	(S) .85
(P) .72	(P) .11	(P) .50	(P) .92

### Key Feature Test

Im Key Feature Tests war das durchschnittliche Testergebnis 66 %; 60 % bei Studierenden im Klinischen Studienabschnitt und 74 % bei Studierenden im Praktischen Jahr (s. Tabelle 11).

Tabelle 11

*Auswertung der Testergebnisse im Key Feature Test*

<b>M</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>
.66	.15	.13	.91
(S) .60	(S) .15	(S) .13	(S) .84
(P) .74	(P) .12	(P) .41	(P) .91

## Simulationen

In den Simulationen war das durchschnittliche Testergebnis 40 %; 39 % bei Studierenden im Klinischen Studienabschnitt und 42 % bei Studierenden im Praktischen Jahr (s. Tabelle 12).

Tabelle 12

*Auswertung der Testergebnisse in den Simulationen*

<b>M</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>
.40	.15	.00	.70
(S) .39	(S) .16	(S) .00	(S) .68
(P) .42	(P) .14	(P) .15	(P) .70

Die Auswertung der einzelnen Patientenfälle ist in den Tabellen 13 bis 17 dargestellt. Die Lage- und Streumaße beziehen sich auf die Testergebnisse, die infolge korrekter Diagnosen weiter ausgewertet wurden.

*Fall Aimüller:*

Im Fall Aimüller wurde in 74 %, von 77 % der Studierenden im Klinischen Studienabschnitt und 72 % der Studierenden im Praktischen Jahr, eine korrekte Diagnose angegeben. In 56 % führte Live-Simulation, in 44 % Video-Simulation zu einer korrekten Diagnose. Das Testergebnis betrug  $M = 1.99$  Punkte ( $SD = 0.48$ ),  $M = 2.16$  ( $SD = 0.38$ ) nach Live-Simulation,  $M = 1.77$  ( $SD = 0.52$ ) nach Video-Simulation. Studierende im Klinischen Studienabschnitt erreichten  $M = 2.01$  ( $SD = 0.49$ ) Punkte,  $M = 2.24$  ( $SD = 0.30$ ) nach Live-Simulation,  $M = 1.73$  ( $SD = 0.55$ ) nach Video-Simulation. Studierende im Praktischen Jahr erreichten  $M = 1.96$  Punkte ( $SD = 0.48$ ),  $M = 2.08$  ( $SD = 0.45$ ) nach Live-Simulation,  $M = 1.82$  ( $SD = 0.50$ ) nach Video-Simulation. Bei der Begründung wurden  $M = 0.29$  Punkte ( $SD = 0.13$ ),  $M = 0.26$  ( $SD = 0.12$ ) von Studierenden im Klinischen Studienabschnitt,  $M = 0.32$  ( $SD = 0.13$ ) von Studierenden im Praktischen Jahr erreicht. Wurde die Diagnose mit 2 Punkten bewertet (70 %), wurden bei der Begründung  $M = 0.29$  Punkte ( $SD = 0.13$ ) erreicht. Die weiteren Ergebnisse der Auswertung sind in Tabelle 13 dargestellt.

Tabelle 13  
*Auswertung der Testergebnisse im Fall Aimüller*

<b>Aimüller</b>	<b>n</b>		<b>M</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>
Punktevergabe	64 (74 %)		1.99	0.48	1.10	2.50
Diagnose	(S) 36		(S) 2.01	(S) 0.49	(S) 1.10	(S) 2.50
	(P) 28		(P) 1.96	(P) 0.48	(P) 1.20	(P) 2.50
Diagnose	45 (70 %)	Begründung	0.29	0.13	0.10	0.50
2 Punkte	(S) 27		(S) 0.28	(S) 0.12	(S) 0.10	(S) 0.50
	(P) 18		(P) 0.31	(P) 0.13	(P) 0.10	(P) 0.50
Diagnose	19 (30 %)	Begründung	0.28	0.14	0.10	0.50
1 Punkt	(S) 9		(S) 0.20	(S) 0.11	(S) 0.10	(S) 0.40
	(P) 10		(P) 0.35	(P) 0.12	(P) 0.20	(P) 0.50
Begründung			0.29	0.13	0.10	0.50
			(S) 0.26	(S) 0.12	(S) 0.10	(S) 0.50
			(P) 0.32	(P) 0.13	(P) 0.10	(P) 0.50
<b>Live-Simulation</b>	36 (56 %)		2.16	0.38	1.10	2.50
	(S) 20		(S) 2.24	(S) 0.30	(S) 1.10	(S) 2.50
	(P) 16		(P) 2.08	(P) 0.45	(P) 1.20	(P) 2.50
Diagnose	31	Begründung	0.30	0.13	0.10	0.50
2 Punkte	(S) 19		(S) 0.30	(S) 0.13	(S) 0.10	(S) 0.50
	(P) 12		(P) 0.32	(P) 0.13	(P) 0.10	(P) 0.50
Diagnose	5	Begründung	0.30	0.16	0.10	0.50
1 Punkt	(S) 1		(S) 0.10	(S) 0.00	(S) 0.10	(S) 0.10
	(P) 4		(P) 0.35	(P) 0.13	(P) 0.20	(P) 0.50
Begründung			0.30	0.13	0.10	0.50
			(S) 0.29	(S) 0.13	(S) 0.10	(S) 0.50
			(P) 0.33	(P) 0.13	(P) 0.10	(P) 0.50
<b>Video-Simulation</b>	28 (44 %)		1.77	0.52	1.10	2.50
	(S) 16		(S) 1.73	(S) 0.55	(S) 1.10	(S) 2.40
	(P) 12		(P) 1.82	(P) 0.50	(P) 1.20	(P) 2.50
Diagnose	14	Begründung	0.26	0.12	0.10	0.50
2 Punkte	(S) 8		(S) 0.25	(S) 0.11	(S) 0.10	(S) 0.40
	(P) 6		(P) 0.28	(P) 0.15	(P) 0.10	(P) 0.50
Diagnose	14	Begründung	0.27	0.13	0.10	0.50
1 Punkt	(S) 8		(S) 0.21	(S) 0.11	(S) 0.10	(S) 0.40
	(P) 6		(P) 0.35	(P) 0.12	(P) 0.20	(P) 0.50
Begründung			0.27	0.12	0.10	0.50
			(S) 0.23	(S) 0.11	(S) 0.10	(S) 0.40
			(P) 0.32	(P) 0.13	(P) 0.10	(P) 0.50

*Fall Albrecht:*

Im Fall Albrecht wurde in 66 %, von 68 % der Studierenden im Klinischen Studienabschnitt und 64 % der Studierenden im Praktischen Jahr, eine korrekte Diagnose angegeben. In 61 % führte Live-Simulation, in 39 % Video-Simulation zu einer korrekten Diagnose. Das Testergebnis betrug  $M = 1.56$  Punkte ( $SD = 0.37$ ),  $M = 1.60$  ( $SD = 0.39$ ) nach Live-Simulation,  $M = 1.49$  ( $SD = 0.33$ ) nach Video-Simulation. Studierende im Klinischen Studienabschnitt erreichten  $M = 1.53$  ( $SD = 0.30$ ) Punkte,  $M = 1.52$  ( $SD = 0.33$ ) nach Live-Simulation,  $M = 1.54$  ( $SD = 0.28$ ) nach Video-Simulation. Studierende im Praktischen Jahr erreichten  $M = 1.60$  Punkte ( $SD = 0.44$ ),  $M = 1.70$  ( $SD = 0.44$ ) nach Live-Simulation,  $M = 1.42$  ( $SD = 0.41$ ) nach Video-Simulation. Bei der Begründung wurden  $M = 0.43$  Punkte ( $SD = 0.16$ ),  $M = 0.43$  ( $SD = 0.13$ ) von Studierenden im Klinischen Studienabschnitt,  $M = 0.44$  ( $SD = 0.19$ ) von Studierenden im Praktischen Jahr erreicht. Wurde die Diagnose mit 2 Punkten bewertet (12 %), wurden bei der Begründung  $M = 0.45$  Punkte ( $SD = 0.14$ ) erreicht. Die weiteren Ergebnisse der Auswertung sind in Tabelle 14 dargestellt.

Tabelle 14  
Auswertung der Testergebnisse im Fall Albrecht

<b>Albrecht</b>	<b>n</b>		<b>M</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>
Punktevergabe	57 (66 %)		1.56	0.37	1.00	2.63
Diagnose	(S) 32		(S) 1.53	(S) 0.30	(S) 1.25	(S) 2.50
	(P) 25		(P) 1.60	(P) 0.44	(P) 1.00	(P) 2.63
Diagnose	7 (12 %)	Begründung	0.45	0.14	0.25	0.63
2 Punkte	(S) 3		(S) 0.38	(S) 0.13	(S) 0.25	(S) 0.50
	(P) 4		(P) 0.50	(P) 0.14	(P) 0.38	(P) 0.63
Diagnose	50 (88 %)	Begründung	0.43	0.16	0.00	0.75
1 Punkt	(S) 29		(S) 0.44	(S) 0.13	(S) 0.25	(S) 0.63
	(P) 21		(P) 0.42	(P) 0.20	(P) 0.00	(P) 0.75
Begründung			0.43	0.16	0.00	0.75
			(S) 0.43	(S) 0.13	(S) 0.25	(S) 0.63
			(P) 0.44	(P) 0.19	(P) 0.00	(P) 0.75
<b>Live-Simulation</b>	35 (61 %)		1.60	0.39	1.25	2.63
	(S) 19		(S) 1.52	(S) 0.33	(S) 1.25	(S) 2.50
	(P) 16		(P) 1.70	(P) 0.44	(P) 1.25	(P) 2.63
Diagnose	5	Begründung	0.48	0.16	0.25	0.63
2 Punkte	(S) 2		(S) 0.38	(S) 0.18	(S) 0.25	(S) 0.50
	(P) 3		(P) 0.54	(P) 0.14	(P) 0.38	(P) 0.63
Diagnose	30	Begründung	0.45	0.14	0.25	0.75
1 Punkt	(S) 17		(S) 0.42	(S) 0.13	(S) 0.25	(S) 0.63
	(P) 13		(P) 0.50	(P) 0.14	(P) 0.25	(P) 0.75
Begründung			0.46	0.14	0.25	0.75
			(S) 0.41	(S) 0.13	(S) 0.25	(S) 0.63
			(P) 0.51	(P) 0.14	(P) 0.25	(P) 0.75
<b>Video-Simulation</b>	22 (39 %)		1.49	0.33	1.00	2.38
	(S) 13		(S) 1.54	(S) 0.28	(S) 1.25	(S) 2.38
	(P) 9		(P) 1.42	(P) 0.41	(P) 1.00	(P) 2.38
Diagnose	2	Begründung	0.38	0.00	0.38	0.38
2 Punkte	(S) 1		(S) 0.38	(S) 0.00	(S) 0.38	(S) 0.38
	(P) 1		(P) 0.38	(P) 0.00	(P) 0.38	(P) 0.38
Diagnose	20	Begründung	0.40	0.18	0.00	0.63
1 Punkt	(S) 12		(S) 0.47	(S) 0.12	(S) 0.25	(S) 0.63
	(P) 8		(P) 0.30	(P) 0.21	(P) 0.00	(P) 0.50
Begründung			0.40	0.17	0.00	0.63
			(S) 0.46	(S) 0.12	(S) 0.25	(S) 0.63
			(P) 0.31	(P) 0.20	(P) 0.00	(P) 0.50



*Fall Klein:*

Im Fall Klein wurde in 74 %, von 70 % der Studierenden im Klinischen Studienabschnitt und 79 % der Studierenden im Praktischen Jahr, eine korrekte Diagnose angegeben. In 58 % führte Live-Simulation, in 42 % Video-Simulation zu einer korrekten Diagnose. Das Testergebnis betrug  $M = 2.44$  Punkte ( $SD = 0.37$ ),  $M = 2.48$  ( $SD = 0.40$ ) nach Live-Simulation,  $M = 2.39$  ( $SD = 0.32$ ) nach Video-Simulation. Studierende im Klinischen Studienabschnitt erreichten  $M = 2.52$  ( $SD = 0.31$ ) Punkte,  $M = 2.60$  ( $SD = 0.28$ ) nach Live-Simulation,  $M = 2.40$  ( $SD = 0.32$ ) nach Video-Simulation. Studierende im Praktischen Jahr erreichten  $M = 2.37$  Punkte ( $SD = 0.41$ ),  $M = 2.36$  ( $SD = 0.47$ ) nach Live-Simulation,  $M = 2.39$  ( $SD = 0.32$ ) nach Video-Simulation. Bei der Begründung wurden  $M = 0.55$  Punkte ( $SD = 0.18$ ),  $M = 0.58$  ( $SD = 0.17$ ) von Studierenden im Klinischen Studienabschnitt,  $M = 0.53$  ( $SD = 0.20$ ) von Studierenden im Praktischen Jahr erreicht. Wurde die Diagnose mit 2 Punkten bewertet (89 %), wurden bei der Begründung  $M = 0.55$  Punkte ( $SD = 0.19$ ) erreicht. Die weiteren Ergebnisse der Auswertung sind in Tabelle 15 dargestellt.

Tabelle 15  
Auswertung der Testergebnisse im Fall Klein

Klein	n		M	SD	Min	Max
Punktevergabe	64 (74 %)		2.44	0.37	1.40	2.80
Diagnose	(S) 33		(S) 2.52	(S) 0.31	(S) 1.40	(S) 2.80
	(P) 31		(P) 2.37	(P) 0.41	(P) 1.40	(P) 2.80
Diagnose	57 (89 %)	Begründung	0.55	0.19	0.00	0.80
2 Punkte	(S) 31		(S) 0.58	(S) 0.17	(S) 0.20	(S) 0.80
	(P) 26		(P) 0.52	(P) 0.20	(P) 0.00	(P) 0.80
Diagnose	7 (11 %)	Begründung	0.54	0.15	0.40	0.80
1 Punkt	(S) 2		(S) 0.50	(S) 0.14	(S) 0.40	(S) 0.60
	(P) 5		(P) 0.56	(P) 0.17	(P) 0.40	(P) 0.80
Begründung			0.55	0.18	0.00	0.80
			(S) 0.58	(S) 0.17	(S) 0.20	(S) 0.80
			(P) 0.53	(P) 0.20	(P) 0.00	(P) 0.80
<b>Live-Simulation</b>	37 (58 %)		2.48	0.40	1.40	2.80
	(S)19		(S) 2.60	(S) 0.28	(S) 1.60	(S) 2.80
	(P)18		(P) 2.36	(P) 0.47	(P) 1.40	(P) 2.80
Diagnose	32	Begründung	0.63	0.16	0.40	0.80
2 Punkte	(S) 18		(S)0.66	(S)0.15	(S)0.40	(S)0.80
	(P) 14		(P)0.59	(P)0.17	(P)0.40	(P)0.80
Diagnose	5	Begründung	0.56	0.17	0.40	0.80
1 Punkt	(S) 1		(S)0.60	(S)0.00	(S)0.60	(S)0.60
	(P) 4		(P)0.55	(P)0.19	(P)0.40	(P)0.80
Begründung			0.62	0.16	0.40	0.80
			(S) 0.65	(S) 0.15	(S) 0.40	(S) 0.80
			(P) 0.58	(P) 0.17	(P) 0.40	(P) 0.80
<b>Video-Simulation</b>	27 (42 %)		2.39	0.32	1.40	2.80
	(S)14		(S) 2.40	(S) 0.32	(S) 1.40	(S) 2.80
	(P)13		(P) 2.39	(P) 0.32	(P) 1.60	(P) 2.80
Diagnose	25	Begründung	0.46	0.19	0.00	0.80
2 Punkte	(S) 13		(S) 0.48	(S) 0.15	(S)0.20	(S) 0.80
	(P) 12		(P) 0.45	(P) 0.23	(P)0.00	(P) 0.80
Diagnose	2	Begründung	0.50	0.14	0.40	0.60
1 Punkt	(S) 1		(S)0.40	(S)0.00	(S) 0.40	(S) 0.40
	(P) 1		(P)0.60	(P)0.00	(P) 0.60	(P) 0.60
Begründung			0.47	0.18	0.00	0.80
			(S) 0.47	(S) 0.15	(S) 0.20	(S) 0.80
			(P) 0.46	(P) 0.22	(P) 0.00	(P) 0.80

*Fall Bircher:*

Im Fall Bircher wurde in 36 %, von 38 % der Studierenden im Klinischen Studienabschnitt und 33 % der Studierenden im Praktischen Jahr, eine korrekte Diagnose angegeben. In 68 % führte Live-Simulation, in 32 % Video-Simulation zu einer korrekten Diagnose. Das Testergebnis betrug  $M = 1.40$  Punkte ( $SD = 0.36$ ),  $M = 1.45$  ( $SD = 0.41$ ) nach Live-Simulation,  $M = 1.30$  ( $SD = 0.16$ ) nach Video-Simulation. Studierende im Klinischen Studienabschnitt erreichten  $M = 1.34$  ( $SD = 0.28$ ) Punkte,  $M = 1.38$  ( $SD = 0.31$ ) nach Live-Simulation,  $M = 1.25$  ( $SD = 0.15$ ) nach Video-Simulation. Studierende im Praktischen Jahr erreichten  $M = 1.48$  Punkte ( $SD = 0.44$ ),  $M = 1.56$  ( $SD = 0.55$ ) nach Live-Simulation,  $M = 1.35$  ( $SD = 0.16$ ) nach Video-Simulation. Bei der Begründung wurden  $M = 0.30$  Punkte ( $SD = 0.12$ ),  $M = 0.28$  ( $SD = 0.10$ ) von Studierenden im Klinischen Studienabschnitt,  $M = 0.33$  ( $SD = 0.13$ ) von Studierenden im Praktischen Jahr erreicht. Wurde die Diagnose mit 2 Punkten bewertet (10 %), wurden bei der Begründung  $M = 0.42$  Punkte ( $SD = 0.07$ ) erreicht. Die weiteren Ergebnisse der Auswertung sind in Tabelle 16 dargestellt.

Tabelle 16  
*Auswertung der Testergebnisse im Fall Bircher*

<b>Bircher</b>	<b>n</b>		<b>M</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>
Punktevergabe	31 (36 %)		1.40	0.36	1.13	2.50
Diagnose	(S) 18		(S) 1.34	(S) 0.28	(S) 1.13	(S) 2.38
	(P) 13		(P) 1.48	(P) 0.44	(P) 1.13	(P) 2.50
Diagnose 2 Punkte	3 (10 %)	Begründung	0.42	0.07	0.38	0.50
	(S) 1		(S) 0.38	(S) 0.00	(S) 0.38	(S) 0.38
	(P) 2		(P) 0.44	(P) 0.09	(P) 0.38	(P) 0.50
Diagnose 1 Punkt	28 (90 %)	Begründung	0.29	0.11	0.13	0.63
	(S) 17		(S) 0.28	(S) 0.10	(S) 0.13	(S) 0.50
	(P) 11		(P) 0.31	(P) 0.13	(P) 0.13	(P) 0.63
Begründung			0.30	0.12	0.13	0.63
			(S) 0.28	(S) 0.10	(S) 0.13	(S) 0.50
			(P) 0.33	(P) 0.13	(P) 0.13	(P) 0.63
<b>Live-Simulation</b>	21 (68 %)		1.45	0.41	1.13	2.50
	(S)13		(S) 1.38	(S) 0.31	(S) 1.13	(S) 2.38
	(P) 8		(P) 1.56	(P) 0.55	(P) 1.13	(P) 2.50
Diagnose 2 Punkte	3	Begründung	0.42	0.07	0.38	0.50
	(S) 1		(S) 0.38	(S) 0.00	(S) 0.38	(S) 0.38
	(P) 2		(P) 0.44	(P) 0.09	(P) 0.38	(P) 0.50
Diagnose 1 Punkt	18	Begründung	0.28	0.08	0.13	0.38
	(S) 12		(S) 0.29	(S) 0.08	(S) 0.13	(S) 0.38
	(P) 6		(P) 0.27	(P) 0.09	(P) 0.13	(P) 0.38
Begründung			0.30	0.09	0.13	0.50
			(S) 0.30	(S) 0.08	(S) 0.13	(S) 0.38
			(P) 0.31	(P) 0.12	(P) 0.13	(P) 0.50
<b>Video-Simulation</b>	10 (32 %)		1.30	0.16	1.13	1.63
	(S) 5		(S) 1.25	(S) 0.15	(S) 1.13	(S)1.50
	(P) 5		(P) 1.35	(P) 0.16	(P) 1.25	(P) 1.63
Diagnose 2 Punkte	0	Begründung	-	-	-	-
	(S) 0		(S) -	(S) -	(S) -	(S) -
	(P) 0		(P) -	(P) -	(P) -	(P) -
Diagnose 1 Punkt	10	Begründung	0.30	0.16	0.13	0.63
	(S) 5		(S) 0.25	(S) 0.15	(S)0.13	(S)0.50
	(P) 5		(P) 0.35	(P) 0.16	(P)0.25	(P)0.63
Begründung			0.30	0.16	0.13	0.63
			(S) 0.25	(S) 0.15	(S) 0.13	(S) 0.50
			(P) 0.35	(P) 0.16	(P) 0.25	(P) 0.63

*Fall Lehner:*

Im Fall Lehner wurde in 71 %, von 64 % der Studierenden im Klinischen Studienabschnitt und 80 % der Studierenden im Praktischen Jahr, eine korrekte Diagnose angegeben. In 39 % führte Live-Simulation, in 61 % Video-Simulation zu einer korrekten Diagnose. Das Testergebnis betrug  $M = 1.73$  Punkte ( $SD = 0.53$ ),  $M = 1.71$  ( $SD = 0.56$ ) nach Live-Simulation,  $M = 1.74$  ( $SD = 0.53$ ) nach Video-Simulation. Studierende im Klinischen Studienabschnitt erreichten  $M = 1.58$  ( $SD = 0.47$ ) Punkte,  $M = 1.63$  ( $SD = 0.52$ ) nach Live-Simulation,  $M = 1.56$  ( $SD = 0.45$ ) nach Video-Simulation. Studierende im Praktischen Jahr erreichten  $M = 1.87$  Punkte ( $SD = 0.56$ ),  $M = 1.76$  ( $SD = 0.60$ ) nach Live-Simulation,  $M = 1.97$  ( $SD = 0.53$ ) nach Video-Simulation. Bei der Begründung wurden  $M = 0.29$  Punkte ( $SD = 0.13$ ),  $M = 0.28$  ( $SD = 0.13$ ) von Studierenden im Klinischen Studienabschnitt,  $M = 0.29$  ( $SD = 0.12$ ) von Studierenden im Praktischen Jahr erreicht. Wurde die Diagnose mit 2 Punkten bewertet (44 %), wurden bei der Begründung  $M = 0.31$  Punkte ( $SD = 0.11$ ) erreicht. Die weiteren Ergebnisse der Auswertung sind in Tabelle 17 dargestellt.

Tabelle 17  
*Auswertung der Testergebnisse im Fall Lehner*

<b>Lehner</b>	<b>n</b>		<b>M</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>
Punktevergabe	61 (71 %)		1.73	0.53	1.00	2.50
Diagnose	(S) 30		(S) 1.58	(S) 0.47	(S) 1.10	(S) 2.40
	(P) 31		(P) 1.87	(P) 0.56	(P) 1.00	(P) 2.50
Diagnose 2 Punkte	27 (44 %)	Begründung	0.31	0.11	0.10	0.50
	(S) 9		(S) 0.26	(S) 0.11	(S) 0.10	(S) 0.40
	(P) 18		(P) 0.33	(P) 0.10	(P) 0.20	(P) 0.50
Diagnose 1 Punkt	34 (56 %)	Begründung	0.27	0.14	0.00	0.60
	(S) 21		(S) 0.29	(S) 0.14	(S) 0.10	(S) 0.60
	(P) 13		(P) 0.24	(P) 0.13	(P) 0.00	(P) 0.50
Begründung			0.29	0.13	0.00	0.60
			(S) 0.28	(S) 0.13	(S) 0.10	(S) 0.60
			(P) 0.29	(P) 0.12	(P) 0.00	(P) 0.50
<b>Live-Simulation</b>	24 (39 %)		1.71	0.56	1.00	2.50
	(S)10		(S) 1.63	(S) 0.52	(S) 1.10	(S) 2.30
	(P)14		(P) 1.76	(P) 0.60	(P) 1.00	(P) 2.50
Diagnose 2 Punkte	11	Begründung	0.29	0.10	0.10	0.50
	(S) 4		(S) 0.23	(S) 0.10	(S) 0.10	(S) 0.30
	(P) 7		(P) 0.33	(P) 0.10	(P) 0.20	(P) 0.50
Diagnose 1 Punkt	13	Begründung	0.22	0.11	0.00	0.40
	(S) 6		(S) 0.23	(S) 0.10	(S) 0.10	(S) 0.40
	(P) 7		(P) 0.20	(P) 0.13	(P) 0.00	(P) 0.40
Begründung			0.25	0.11	0.00	0.50
			(S) 0.23	(S) 0.09	(S) 0.10	(S) 0.40
			(P) 0.26	(P) 0.13	(P) 0.00	(P) 0.50
<b>Video-Simulation</b>	37 (61 %)		1.74	0.53	1.10	2.50
	(S) 20		(S) 1.56	(S) 0.45	(S) 1.10	(S) 2.40
	(P) 17		(P) 1.97	(P) 0.53	(P) 1.20	(P) 2.50
Diagnose 2 Punkte	16	Begründung	0.32	0.11	0.10	0.50
	(S) 5		(S) 0.28	(S) 0.13	(S) 0.10	(S) 0.40
	(P) 11		(P) 0.34	(P) 0.10	(P) 0.20	(P) 0.50
Diagnose 1 Punkt	21	Begründung	0.31	0.15	0.10	0.60
	(S) 15		(S) 0.31	(S) 0.16	(S) 0.10	(S) 0.60
	(P) 6		(P) 0.28	(P) 0.13	(P) 0.20	(P) 0.50
Begründung			0.31	0.13	0.10	0.60
			(S) 0.31	(S) 0.15	(S) 0.10	(S) 0.60
			(P) 0.32	(P) 0.11	(P) 0.20	(P) 0.50

## 3.2 Auswertung der Fragestellungen

### 3.2.1 Korrelation Performanz in Wissenstests

Die Performanz im Konzeptuellen Wissenstest und Key Feature Test korrelierte positiv mit  $r = .73$  ( $p < .01$ ),  $r = .67$  ( $p < .01$ ) bei Studierenden im Klinischen Studienabschnitt sowie  $r = .66$  ( $p < .01$ ) bei Studierenden im Praktischen Jahr (s. Anhang 8, Abbildung 8).

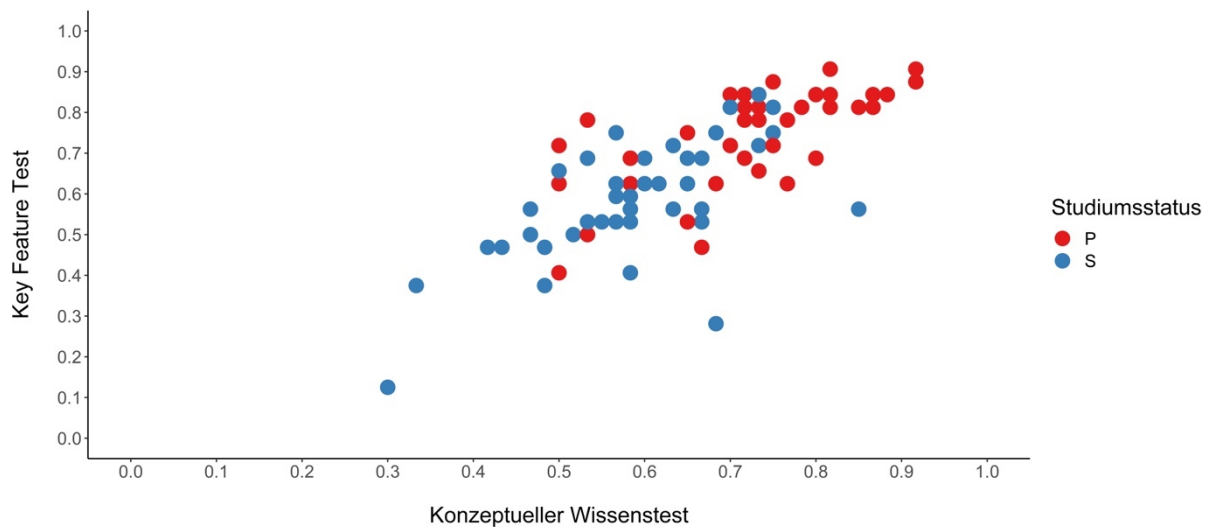


Abbildung 8. Korrelation nach Pearson zwischen der Performanz im Konzeptuellen Wissenstests und Key Feature Test unter Berücksichtigung des Studiumsstatus

### 3.2.2 Korrelation Performanz in Wissenstests und Simulationen

Die Performanz in Wissenstests und Simulationen korrelierte positiv mit  $r = .27$  ( $p = .01$ ),  $r = .30$  ( $p < .01$ ) in Konzeptuellem Wissenstest und Simulationen,  $r = .22$  ( $p = .05$ ) in Key Feature Test und Simulationen (s. Abbildung 9, Tabelle 18).

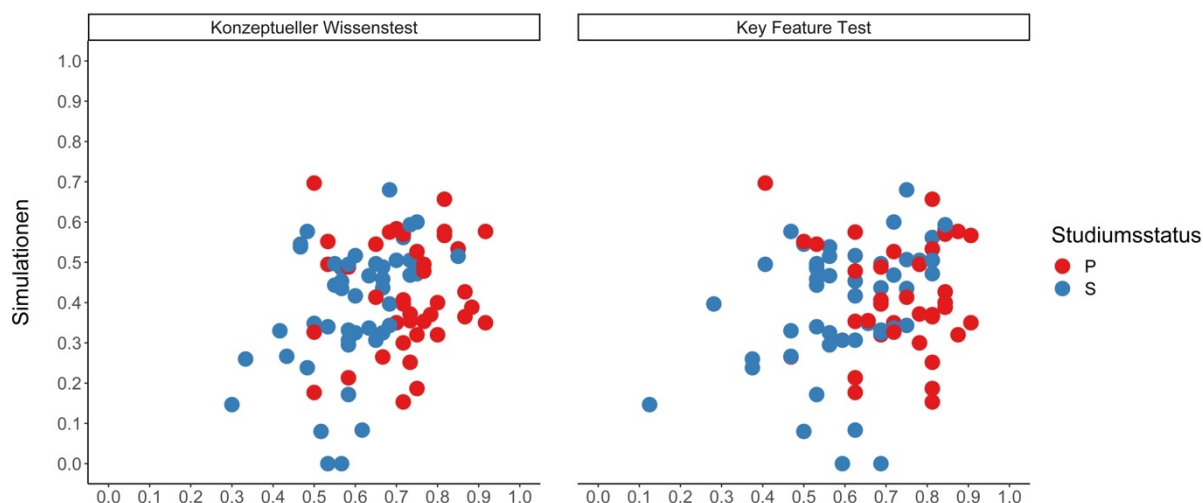


Abbildung 9. Korrelation nach Pearson zwischen der Performanz in Konzeptuellem Wissenstest und Simulationen sowie Key Feature Test und Simulationen unter Berücksichtigung des Studiumsstatus

Die Effektgrößen wurden unabhängig der Gruppe und somit dem Bearbeitungsmodus der Patientenfälle von Studierenden im Klinischen Studienabschnitt hervorgerufen. Bei Studierenden im Praktischen Jahr zeigte sich kein positiver linearer Zusammenhang zwischen den Testergebnissen der Messinstrumente (s. Tabelle 18, s. Abbildung 9).

Tabelle 18

Korrelation nach Pearson zwischen der Performanz in Wissenstests und Simulationen insgesamt und unter Aufteilung nach Gruppe

	Simulationen (Gruppe 1 + Gruppe 2)	Gruppe 1 (n = 40)	Gruppe 2 (n = 46)
Wissenstests	.27* (p = .01)	.28 (p = .08)	.25 (p = .09)
(KW-Test + KF-Test)	(S) .42** (p < .01)	(S) .39 (p = .08)	(S) .45* (p = .02)
	(P) .01 (p = .94)	(P) .09 (p = .71)	(P) -.10 (p = .68)
Konzeptueller	.30* (p < .01)	.29 (p = .07)	.29* (p = .05)
Wissenstest	(S) .45** (p < .01)	(S) .39 (p = .07)	(S) .47* (p = .02)
	(P) .07 (p = .69)	(P) .16 (p = .54)	(P) .01 (p = .98)
Key Feature Test	.22* (p = .05)	.25 (p = .12)	.19 (p = .22)
	(S) .34* (p = .02)	(S) .36 (p = .10)	(S) .34 (p = .10)
	(P) -.05 (p = .78)	(P) .02 (p = .93)	(P) -.16 (p = .49)

\*\* Die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant.

\* Die Korrelation ist auf dem Niveau von 0.05 (2-seitig) signifikant.



In Gruppe 1, der Anamneseerhebung der Fälle Bircher und Lehner in Live-Simulation, korrelierten die Testergebnisse aus Wissenstests und Live-Simulation positiv mit  $r = .44$  ( $p < .01$ ) (s. Tabelle 19).

Tabelle 19

*Gruppe 1: Korrelation nach Pearson zwischen der Performanz in Wissenstests und Simulationen insgesamt und unter Aufteilung nach Simulationsart*

<b>n = 40 (S) = 22, (P) = 18</b>	<b>Simulationen</b>	<b>Live-Simulation</b>	<b>Video-Simulation</b>
Wissenstests	.28 (p = .08)	.44** (p < .01)	.06 (p = .70)
(KW-Test + KF-Test)	(S) .39 (p = .08)	(S) .51* (p = .02)	(S) .19 (p = .39)
	(P) .09 (p = .71)	(P) .26 (p = .31)	(P) -.06 (p = .81)
Konzeptueller Wissenstest	.29 (p = .07)	.44** (p < .01)	.07 (p = .65)
	(S) .39 (p = .07)	(S) .45* (p = .04)	(S) .24 (p = .29)
	(P) .16 (p = .54)	(P) .37 (p = .14)	(P) -.06 (p = .81)
Key Feature Test	.25 (p = .12)	.40* (p = .01)	.05 (p = .76)
	(S) .36 (p = .10)	(S) .52* (p = 0.01)	(S) .15 (p = .52)
	(P) .02 (p = .93)	(P) .12 (p = .64)	(P) -.05 (p = .84)
Live-Simulation			.17 (p = .30)
			(S) .36 (p = .10)
			(P) -.07 (p = .78)

\*\* Die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant.

\* Die Korrelation ist auf dem Niveau von 0.05 (2-seitig) signifikant.

In Gruppe 2, der Anamneseerhebung der Fälle Aimüller, Albrecht und Klein in Live-Simulation, korrelierten die Testergebnisse aus Wissenstests und Video-Simulation positiv mit  $r = .42$  ( $p < .01$ ) (s. Tabelle 20).

Tabelle 20

Gruppe 2: Korrelation nach Pearson zwischen der Performanz in Wissenstests und Simulationen insgesamt und unter Aufteilung nach Simulationsart

<b>n = 46 (S) = 25, (P) = 21</b>	<b>Simulationen</b>	<b>Live-Simulation</b>	<b>Video-Simulation</b>
Wissenstests (KW-Test + KF-Test)	.25 ( $p = .09$ ) (S) .45* ( $p = .02$ ) (P) - .10 ( $p = .68$ )	.04 ( $p = .78$ ) (S) .21 ( $p = .31$ ) (P) - .20 ( $p = .39$ )	.42** ( $p < .01$ ) (S) .64** ( $p < .01$ ) (P) .10 ( $p = .68$ )
Konzeptueller Wissenstest	.29* ( $p = .05$ ) (S) .47* ( $p = .02$ ) (P) .01 ( $p = .98$ )	.13 ( $p = .41$ ) (S) .33 ( $p = .11$ ) (P) - .13 ( $p = .59$ )	.36* ( $p = .01$ ) (S) .45* ( $p = .02$ ) (P) .17 ( $p = .46$ )
Key Feature Test	.19 ( $p = .22$ ) (S) .34 ( $p = .10$ ) (P) - .16 ( $p = .49$ )	- .02 ( $p = .89$ ) (S) .09 ( $p = .67$ ) (P) - .22 ( $p = .34$ )	.39** ( $p < .01$ ) (S) .62** ( $p < .01$ ) (P) .02 ( $p = .93$ )
Live-Simulation			.08 ( $p = .61$ ) (S) .16 ( $p = .45$ ) (P) .01 ( $p = .97$ )

\*\* Die Korrelation ist auf dem Niveau von 0.01 (2-seitig) signifikant.

\* Die Korrelation ist auf dem Niveau von 0.05 (2-seitig) signifikant.

Performanzen in Wissenstests und Simulationen korrelierten somit nicht aufgrund einer Simulationsart positiv, sondern fallspezifisch bei Bearbeitung der Fälle Bircher und Lehner. Die Testergebnisse zwischen beiden Simulationsarten korrelierten in Gruppe 1 mit  $r = .17$  ( $p = .30$ ) sowie  $r = .08$  ( $p = .61$ ) in Gruppe 2.

### 3.2.3 Assoziationsfaktoren von Performanz in Simulationen

#### a) Art der Simulation

Live-Simulation für das Anamnesegespräch war in den Fällen Aimüller, Albrecht, Klein und Bircher mit einer höheren Wahrscheinlichkeit für korrekte Diagnosen assoziiert, Video-Simulation im Fall Lehner, unabhängig des Studiumsstatus. Die Differenz der Wahrscheinlichkeiten für korrekte Diagnosen in Live- oder Video-Simulation lag zwischen 8.26 % im Fall Aimüller und 30.76 % im Fall Bircher. Die Fälle Albrecht, Bircher und Lehner zeigten sich in Bezug auf die Simulationsart am sensibelsten (s. Tabelle 21).

Tabelle 21

*Wahrscheinlichkeiten für korrekte Diagnosen in den Patientenfällen der Simulationen unter Aufteilung nach Simulationsart*

		<b>Aimüller</b>	<b>Albrecht</b>	<b>Klein</b>	<b>Bircher</b>	<b>Lehner</b>
<b>Wahrscheinlichkeit für korrekte Diagnose (%)</b>	<b>Video</b>	70.00	55.00	67.50	21.74	80.43
		(S) 72.73	(S) 59.09	(S) 63.63	(S) 20.00	(S) 80.00
		(P) 66.67	(P) 50.00	(P) 72.22	(P) 23.81	(P) 80.95
	<b>Live</b>	78.26	76.09	80.43	52.50	60.00
		(S) 80.00	(S) 76.00	(S) 76.00	(S) 59.09	(S) 45.45
		(P) 76.19	(P) 76.19	(P) 85.71	(P) 44.44	(P) 77.78

Kongruent zu den Ergebnissen der Wahrscheinlichkeiten, resultierten aus der Anamneseerhebung in Live-Simulation mit Ausnahme im Fall Lehner mehr korrekte Diagnosen und bessere Testergebnisse (s. 3.1.2: Tabellen 13-17).

#### b) Effekt der Reihenfolge

Die Startbedingung Live-Simulation (A) war mit einer höheren Wahrscheinlichkeit für korrekte Diagnosen in Video-Simulation in den Fällen Aimüller und Albrecht assoziiert. Die Chance für eine korrekte Diagnose nach Video-Simulation war hierdurch bis um den Faktor 4 (95 % -KI: [0.89,18.03], Aimüller) erhöht. Die Startbedingung Video-Simulation (B) war mit einer höheren Wahrscheinlichkeit für korrekte Diagnosen in Live-Simulation in den Fällen Aimüller, Albrecht und Klein assoziiert. Die Chance für eine korrekte Diagnose nach Live-Simulation war hierdurch bis um den Faktor 4 (95 % -KI: [0.90,17.74], Albrecht) erhöht (s. Tabelle 22). Die Auswertung unter Aufteilung nach Studiumsstatus ist als Anhang beigelegt (s. Anhang 9, s. 4.2.2).

Tabelle 22

*Effekt der Reihenfolge auf die Wahrscheinlichkeiten für korrekte Diagnosen in den Patientenfällen der Simulationen unter Aufteilung nach Simulationsart und Startbedingung*

		Aimüller	Albrecht	Klein	Bircher	Lehner
<b>Wahrscheinlichkeit<sup>1</sup> für korrekte Diagnose (%)</b>	<b>Video/Start Live</b>	84.21	63.16	63.16	18.18	77.27
	Video/Start Video	57.14	47.62	71.43	25.00	83.33
	<b>Live/Start Video</b>	79.17	87.50	87.50	47.62	57.14
	Live/Start Live	77.27	63.64	72.73	57.89	63.16
<b>Odds Ratio</b> [95 % -Konfidenzintervall]		4.00	1.89	0.69	0.67	0.68
	<b>Video/Start Live</b>	[0.89, 18.03]	[0.53, 6.69]	[0.18, 2.59]	[0.16, 2.77]	[0.16, 2.94]
		1.12	4.00	2.62	0.66	0.78
	<b>Live/Start Video</b>	[0.28, 4.54]	[0.90, 17.74]	[0.57, 12.13]	[0.19, 2.31]	[0.22, 2.77]

<sup>1</sup>Bearbeitete Simulationsart/ Startsimulationsart

Die Startbedingung Live-Simulation (A) zeigte mehr korrekte Diagnosen (1A = 40 vs. 1B = 37) und bessere Testergebnisse (1A = 5.68 vs. 1B = 5.63) in Video-Simulation in der Gruppierung Aimüller, Albrecht, Klein. In der Gruppierung Bircher, Lehner wurden weniger korrekte Diagnosen (2A = 21 vs. 2B = 26) und schlechtere Testergebnisse (2A = 2.94 vs. 2B = 3.12) in Video-Simulation erzielt. Die Startbedingung Video-Simulation (B) führte zu mehr korrekten Diagnosen (2B = 61 vs. 2A = 47), und schlechteren Testergebnissen (2B = 6.14 vs. 2A = 6.36) in Live-Simulation in der Gruppierung Aimüller, Albrecht, Klein. In der Gruppierung Bircher, Lehner wurden weniger korrekte Diagnosen (1B = 22 vs. 1A = 23) und schlechtere Testergebnisse (1B = 3.13 vs. 1A = 3.19) in Live Simulation erzielt (s. Tabelle 23).

Tabelle 23

*Performanz in den Patientenfällen der Simulationen unter Aufteilung nach Gruppe und Startbedingung*

Gruppe	Diagnose	Aimüller	Albrecht	Klein	Bircher	Lehner
<b>1A</b>	<b>2 Punkte</b>	<b>9 (47 %)</b>	-	<b>12 (63 %)</b>	<b>1 (5 %)</b>	<b>6 (32 %)</b>
<b>(n = 19)</b>	<b>(n)</b>	(S) 5	(S) -	(S) 6	(S) -	(S) 2
(S) 10		(P) 4	(P) -	(P) 6	(P) 1	(P) 4
(P) 9	<b>1 Punkt</b>	<b>7 (37 %)</b>	<b>12 (63 %)</b>	-	<b>10 (53 %)</b>	<b>6 (32 %)</b>
	<b>(n)</b>	(S) 3	(S) 7	(S) -	(S) 6	(S) 2
		(P) 4	(P) 5	(P) -	(P) 4	(P) 4
	<b>Test- ergebnis</b>	<b>1.85 (0.53)</b>	<b>1.36 (0.16)</b>	<b>2.47 (0.18)</b>	<b>1.41 (0.37)</b>	<b>1.78 (0.60)</b>
	<b>(M/SD)</b>	(S) 1.86 (0.56)	(S) 1.41 (0.12)	(S) 2.50 (0.17)	(S) 1.33 (0.07)	(S) 1.75 (0.64)
		(P) 1.84 (0.54)	(P) 1.30 (0.21)	(P) 2.43 (0.20)	(P) 1.50 (0.57)	(P) 1.79 (0.62)

Gruppe	Diagnose	Aimüller	Albrecht	Klein	Bircher	Lehner
<b>1B</b>	<b>2 Punkte</b>	<b>5 (24 %)</b>	<b>2 (10 %)</b>	<b>13 (62 %)</b>	<b>2 (10 %)</b>	<b>5 (24 %)</b>
<b>(n = 21)</b>	<b>(n)</b>	<b>(S) 3</b>	<b>(S) 1</b>	<b>(S) 7</b>	<b>(S) 1</b>	<b>(S) 2</b>
<b>(S) 12</b>		<b>(P) 2</b>	<b>(P) 1</b>	<b>(P) 6</b>	<b>(P) 1</b>	<b>(P) 3</b>
<b>(P) 9</b>	<b>1 Punkt</b>	<b>7 (33 %)</b>	<b>8 (38 %)</b>	<b>2 (10 %)</b>	<b>8 (38 %)</b>	<b>7 (33 %)</b>
	<b>(n)</b>	<b>(S) 5</b>	<b>(S) 5</b>	<b>(S) 1</b>	<b>(S) 6</b>	<b>(S) 4</b>
		<b>(P) 2</b>	<b>(P) 3</b>	<b>(P) 1</b>	<b>(P) 2</b>	<b>(P) 3</b>
	<b>Test- ergebnis (M/SD)</b>	<b>1.66 (0.51)</b> (S) 1.60 (0.54) (P) 1.78 (0.51)	<b>1.64 (0.43)</b> (S) 1.69 (0.34) (P) 1.56 (0.58)	<b>2.33 (0.39)</b> (S) 2.33 (0.40) (P) 2.34 (0.41)	<b>1.49 (0.47)</b> (S) 1.41 (0.43) (P) 1.67 (0.62)	<b>1.64 (0.54)</b> (S) 1.55 (0.47) (P) 1.73 (0.62)
<b>2A</b>	<b>2 Punkte</b>	<b>16 (72 %)</b>	<b>2 (9 %)</b>	<b>15 (68 %)</b>	<b>-</b>	<b>6 (27 %)</b>
<b>(n = 22)</b>	<b>(n)</b>	<b>(S) 10</b>	<b>(S) -</b>	<b>(S) 8</b>	<b>(S) -</b>	<b>(S) 1</b>
<b>(S) 12</b>		<b>(P) 6</b>	<b>(P) 2</b>	<b>(P) 7</b>	<b>(P) -</b>	<b>(P) 5</b>
<b>(P) 10</b>	<b>1 Punkt</b>	<b>1 (5 %)</b>	<b>12 (55 %)</b>	<b>1 (5 %)</b>	<b>4 (18 %)</b>	<b>11 (50 %)</b>
	<b>(n)</b>	<b>(S) -</b>	<b>(S) 8</b>	<b>(S) -</b>	<b>(S) 4</b>	<b>(S) 8</b>
		<b>(P) 1</b>	<b>(P) 4</b>	<b>(P) 1</b>	<b>(P) -</b>	<b>(P) 3</b>
	<b>Test- ergebnis (M/SD)</b>	<b>2.22 (0.29)</b> (S) 2.27 (0.12) (P) 2.16 (0.44)	<b>1.58 (0.41)</b> (S) 1.41 (0.15) (P) 1.81 (0.55)	<b>2.56 (0.34)</b> (S) 2.63 (0.17) (P) 2.50 (0.47)	<b>1.25 (0.18)</b> (S) 1.25 (0.18) (P) -	<b>1.69 (0.48)</b> (S) 1.46 (0.31) (P) 1.96 (0.50)
<b>2B</b>	<b>2 Punkte</b>	<b>15 (63 %)</b>	<b>3 (13 %)</b>	<b>17 (71 %)</b>	<b>-</b>	<b>10 (42 %)</b>
<b>(n = 24)</b>	<b>(n)</b>	<b>(S) 9</b>	<b>(S) 2</b>	<b>(S) 10</b>	<b>(S) -</b>	<b>(S) 4</b>
<b>(S) 13</b>		<b>(P) 6</b>	<b>(P) 1</b>	<b>(P) 7</b>	<b>(P) -</b>	<b>(P) 6</b>
<b>(P) 11</b>	<b>1 Punkt</b>	<b>4 (17 %)</b>	<b>18 (75 %)</b>	<b>4 (17 %)</b>	<b>6 (25 %)</b>	<b>10 (42 %)</b>
	<b>(n)</b>	<b>(S) 1</b>	<b>(S) 9</b>	<b>(S) 1</b>	<b>(S) 1</b>	<b>(S) 7</b>
		<b>(P) 3</b>	<b>(P) 9</b>	<b>(P) 3</b>	<b>(P) 5</b>	<b>(P) 3</b>
	<b>Test- ergebnis (M/SD)</b>	<b>2.11 (0.44)</b> (S) 2.20 (0.41) (P) 2.01 (0.47)	<b>1.61 (0.38)</b> (S) 1.60 (0.40) (P) 1.63 (0.38)	<b>2.42 (0.44)</b> (S) 2.58 (0.35) (P) 2.24 (0.47)	<b>1.33 (0.15)</b> (S) 1.25 (0.00) (P) 1.35 (0.16)	<b>1.79 (0.57)</b> (S) 1.64 (0.55) (P) 1.97 (0.58)

## 4 Diskussion

### 4.1 Zusammenfassung der Ergebnisse

#### 4.1.1 Zusammenfassung der deskriptiven Analysen

##### 4.1.1.1 Item- und Skalenanalyse der Wissenstests

Der Konzeptuelle Wissenstest wurde für die Auswertung um 10 Fragen mit einer Trennschärfe  $< .1$  gekürzt. Die Schwierigkeit der 30 verbliebenen Items lag zwischen  $.10$  und  $.95$ , bei 11 war die Trennschärfe  $> .3$ . Die Reliabilität unter Berechnung der internen Konsistenz erreichte  $\alpha = .76$ .

Im Key Feature Test resultierten aus der Fallanalyse Schwierigkeiten zwischen  $.33$  und  $.93$  sowie Trennschärfen  $> .3$  in 9 von 10 Fällen. 8 Fragen mit einer Trennschärfe  $< .1$  wurden für die Auswertung ausgeschlossen. Die Schwierigkeit der 32 verbliebenen Items lag zwischen  $.13$  und  $.99$ , bei 16 war die Trennschärfe  $> .3$ . Die Reliabilität unter Berechnung der internen Konsistenz erreichte  $\alpha = .81$ .

##### 4.1.1.2 Testergebnisse der Messinstrumente

Sowohl im Konzeptuellen Wissenstest als auch im Key Feature Test war das durchschnittliche Testergebnis 66 %. In den Simulationen wurden im Durchschnitt 40 % der möglichen Punkte erreicht. In allen 3 Messinstrumenten erzielten Studierende im Praktischen Jahr bessere Testergebnisse als Studierende im Klinischen Studienabschnitt.

Die Auswertung der Patientenfälle zeigte eine starke fallspezifische Variation der Testergebnisse: der Anteil korrekter Diagnosen lag zwischen 36 % (Bircher) und 74 % (Aimüller, Klein). Diese wurden mehr durch Anamnese in Live-Simulation als nach Video-Simulationen gestellt (ausgenommen Lehner). Im Vergleich der einzelnen Fälle war ein positiver Zusammenhang zwischen der Anzahl korrekter Diagnosen und dem durchschnittlichen Testergebnis erkennbar. Unabhängig der Simulationsart erreichten Studierende im Klinischen Studienabschnitt bessere Testergebnisse im Fall Klein, Studierende im Praktischen Jahr in den Fällen Bircher und Lehner. Im Fall Aimüller zeigten Studierende im Klinischen Studienabschnitt bessere Testergebnisse insgesamt und nach Live-Simulation, nach Video-Simulation Studierende im Praktischen Jahr. Im Fall Albrecht erzielten Studierende im Praktischen Jahr bessere Testergebnisse insge-

samt und nach Live-Simulation, nach Video-Simulation Studierende im Klinischen Studienabschnitt. Die Streuung der Testergebnisse lag zwischen  $SD = 0.36$  Punkten (Bircher) und  $SD = 0.53$  Punkten (Lehner) und war nach Live-Simulation stärker als nach Video-Simulation (ausgenommen Aimüller). Die Testergebnisse von Studierenden im Praktischen Jahr zeigten in allen Fällen eine stärkere Streuung nach Live-Simulation, ebenso nach Video-Simulation mit Ausnahme der Fälle Aimüller und Klein. In diesen beiden Fällen wurden von Studierenden im Klinischen Studienabschnitt mehr *2 Punkte* Diagnosen gestellt, in den Fällen Albrecht, Bircher und Lehner von Studierenden im Praktischen Jahr. *2 Punkte* Diagnosen zeigten in allen Fällen eine bessere Begründungsqualität als *1 Punkt* Diagnosen. Studierende im Praktischen Jahr erzielten bessere Ergebnisse bei der Begründung insgesamt, bei *2 Punkte* Diagnosen (ausgenommen Klein) sowie *1 Punkt* Diagnosen (ausgenommen Albrecht) als Studierende im Klinischen Studienabschnitt.

#### 4.1.2 Zusammenfassung der Fragestellungen

Die Performanzen im Konzeptuellen Wissenstest und Key Feature Test korrelierten mit  $r = .73$  ( $p < .01$ ).

Die Performanzen in Wissenstests und Simulationen korrelierten mit  $r = .27$  ( $p = .01$ ). Positive Korrelationen zeigten sich nicht aufgrund einer Simulationsart, sondern fallspezifisch. Die Performanz in Simulationen korrelierte stärker mit der Performanz im Konzeptuellen Wissenstest als im Key Feature Test. Die Aufteilung nach Studiumsstatus zeigte, dass die Effektgrößen durch die Testergebnisse von Studierenden im Klinischen Studienabschnitt hervorgerufen wurden. Bei Studierenden im Praktischen Jahr war kein Zusammenhang zwischen den Testergebnissen der Messinstrumente zu erkennen.

Live-Simulationen für Anamnesegespräche führten, mit einer Ausnahme, zu höheren Wahrscheinlichkeiten für korrekte Diagnosen sowie besseren Testergebnissen als Video-Simulationen. Vorschalten von Simulationen erhöhte fallspezifisch die Wahrscheinlichkeit für korrekte Diagnosen. Das Vorschalten von Live-Simulationen steigerte die Chance für korrekte Diagnosen in Video-Simulation bis um den Faktor 4. Das Vorschalten von Video-Simulationen erhöhte die Chance für korrekte Diagnosen in Live-Simulation ebenso bis um den Faktor 4. Die Konfidenzintervalle der Odds Ratios waren breit. In der deskriptiven Auswertung zeigten sich gruppierungsspezifisch durch das Vorschalten von Live-Simulationen mehr korrekte Diagnosen und bessere Testergebnisse in Video-Simulation, durch das Vorschalten von Video-Simulationen mehr korrekte Diagnosen, jedoch keine besseren Testergebnisse in Live-Simulation.

## 4.2 Diskussion der Ergebnisse

### 4.2.1 Diskussion der deskriptiven Analysen

#### 4.2.1.1 Item- und Skalenanalyse der Wissenstests

In der Literatur werden Frageitems von Wissenstests ab .85 als leicht, bis .49 als schwer und im dazwischen liegenden Bereich als mittel klassifiziert [95]. Unter dieser Referenz zeigte sich eine breite Streuung in beiden Wissenstests. Bortz et al. (2002) empfehlen für die Schwierigkeit einen Referenzbereich von .2 bis .8 [96]. Dieser wurde in beiden Wissenstests nicht eingehalten. Bauer et al. (2011) zeigten, dass sowohl Single Choice Typ A als auch Pick-N Multiple Choice Antwortformate für die Konzeption von Wissenstests verwendbar sind [93]. Der Vergleich der Frageitems im Konzeptuellen Wissenstest bestätigt, dass zwischen Pick-N Multiple Choice Format (s. Tabelle 7 *kursive* Items) und Single Choice Typ A Format keine erheblichen Unterschiede für Schwierigkeit und Trennschärfe hervorgerufen wurden. Beide Antwortformate können somit gleichermaßen zur Messung von klinischen Diagnosekompetenzen bei Medizinstudierenden eingesetzt werden. Als Hinweis darauf, dass Frageitems konsistent Bezug zur gemeinsamen Skala aufweisen, sollte die Trennschärfe eines Items Werte  $> .5$  erreichen [97]. Der Konzeptuelle Wissenstest beinhaltete 11, der Key Feature Test 16 Frageitems mit lediglich einer Trennschärfe  $> .3$ . Die Testergebnisse der Fragen innerhalb einer Skala waren somit oft inhomogen. Ziel war, unter Inkaufnahme von Informationsverlusten, durch Streichung von Frageitems mit Trennschärfen  $< .1$  die Reliabilität und somit die Performanz der einzelnen Skalen zu verbessern. Konsequenterweise ergaben sich daraus Veränderungen bei den übrigen Items. Abnahmen der Trennschärfe erklären sich beispielsweise aus zufälligen Korrelationen mit einem entfernten Item. Die Item-Anzahl innerhalb der Skalen beider Wissenstests variierte und war insgesamt gering, weshalb Reliabilitätsverbesserungen der Gesamtskalen durch Item-Elimination sowie reliable Messungen der theoretischen Konstrukte mit Vorsicht zu interpretieren sind. Für die Reliabilität empfehlen Bortz et al. (2006) Werte von  $> .8$  [97]. Diese konnte von keiner der Skalen in den Wissenstests erreicht werden. Eine Verbesserung der Reliabilität könnte durch das Weglassen schlechter Items, die Erhöhung der Item-Anzahl sowie eine inhaltlich breitere Auswahl an Fragen erzielt werden. Die Reliabilität der Wissenstests insgesamt erreichte hohe Werte von  $\alpha = .76$  im Konzeptuellen Wissenstest und  $\alpha = .81$  im Key Feature Test.

Die Aussage von psychometrischen Variablen wie Test-Reliabilität oder Item-Schwierigkeit sollte aber nicht überschätzt werden, wenn durch den Gesamttest verschiedene Aspekte von Diagnosekompetenzen gemessen werden und somit Validität erreicht werden kann.



#### 4.2.1.2 Testergebnisse der Messinstrumente

In der Auswertung der Messinstrumente zeigte sich der Einfluss von personenspezifischen und externen Faktoren auf Diagnosekompetenzen (s. 1.3.3). Die besseren Testergebnisse von Studierenden im Praktischen Jahr bestätigten den von Boshuizen et al. (2019) dargestellten positiven Einfluss von Expertise auf Diagnosekompetenzen [17]. Die Kohärenz über alle Messinstrumente unterschiedlicher Formate hinweg verdeutlicht, dass die von Mandl et al. (2004) beschriebene kognitive Flexibilität sowie multiple Wissensrepräsentationen als Kennzeichen medizinischer Expertise zu werten sind [20]. Wie Stark et al. (2011) bereits darlegten, war der Umfang der Wissensbasis (mit fortschreitender Studiumsdauer) vorteilhaft in allen Lernumgebungen [48]. Darüber hinaus konnten durch die Aufteilung nach Studiumsstatus, wie von Patel und Groen (1991) beschrieben, verschiedene Stadien medizinischer Expertise (s.1.3.1) erkannt werden [52]. Die Anwendung von theoretischem und praktischem Wissen im täglichen Patientenkontakt während des Praktischen Jahres scheint die von Schmidt und Rikers (2007) aufgezeigte Wissensenkapsulierung zu bestätigen [38], epistemisch-diagnostische Aktivitäten zu trainieren und Diagnosequalität zu fördern.

Die Kohorte der Medizinstudierenden im Klinischen Studienabschnitt setzte sich aus drei Studienjahren zusammen, die Studierenden im Praktischen Jahr befanden sich, aufgrund des Curriculums und der Erhebungstage der Studie, entweder am Anfang oder am Ende des Praktischen Jahres. In den Simulationen waren die Testergebnisse der einzelnen Patientenfälle bei Studierenden im Praktischen Jahr mit einer stärkeren Streuung behaftet als bei Studierenden im Klinischen Studienabschnitt. Anamnesegespräche im täglichen Patientenkontakt während des Praktischen Jahres könnten somit maßgeblich zur Entwicklung von Diagnosekompetenzen beitragen und die stärkere Streuung erklären.

Insgesamt zeigten die Testergebnisse der Simulationen eine starke Fallbezogenheit. Die von Jensen et al. (2019) beschriebene Case Specificity, dass Diagnosekompetenzen nicht unbedingt als generalisierbar zu erachten sind, ist somit zu bejahen [21]. Darüber hinaus war die Case Specificity in der fallspezifischen Häufigkeit von *2 Punkten* bzw. *1 Punkt* bei der Auswertung der Diagnose zu erkennen. Die Qualität der Begründung war mit der Höhe der Punkte der Diagnose positiv assoziiert. Patientenfälle mit typisch klinischen Manifestationen von Atemnot wie Lungenembolie (Aimüller) oder Pneumonie (Klein) wurden am häufigsten korrekt und mit den besten Testergebnissen diagnostiziert. Bei Atemnot in Kombination mit Komorbiditäten oder seltenen, genetischen Ursachen wie Hyperventilation bei Panikattacke (Lehner) oder

Herzinsuffizienz bei HOCM bzw. TAA (Albrecht bzw. Bircher) wurden weniger korrekte Diagnosen sowie schlechtere Testergebnisse erzielt. Diese Patientenfälle, in der weiteren Diskussion als spezifische Diagnosen bezeichnet, wurden seltener mit 2 *Punkten* korrekt diagnostiziert. Die Komplexität eines Patientenfalls als externer Einflussfaktor auf die Messung von klinischen Diagnosekompetenzen war hierbei zu erkennen. Interessanter Weise wurden die Diagnosen Pneumonie und Lungenembolie mit typisch klinischer Manifestation von Studierenden im Praktischen Jahr seltener mit 2 *Punkten* korrekt diagnostiziert. Die Patientenfälle könnten für sie bereits ausreichend saliente Merkmale beinhaltet haben, um darin klinische Muster eines Krankheitsbildes zu erkennen. Im Rahmen der von Mattingly und Fleming (1994) beschriebenen Pattern Recognition [58] könnte die Diagnose folglich mehr intuitiv über das System 1 als hypothetisch-deduktiv (s.1.3.2), und deshalb ungenauer angegeben worden sein. Kürzere Zeitspannen von Beendigung des Anamnesegesprächs bis zur Eingabe der Diagnose könnten diesbezüglich Aufschluss verleihen und anhand der Zeiterfassung über CASUS geprüft werden. Als weiterer Situativer Einflussfaktor auf die Messung von Diagnosekompetenzen zeigte sich die Simulationsart als Umstand der Diagnosesituation. Live-Simulationen führten größtenteils, vermutlich aufgrund der Möglichkeit einer direkten Kommunikation, zu besseren Testergebnissen. Eine inhaltliche Unterstützung durch das Auswahlmenü in Video-Simulationen ist daher eher auszuschließen. Die größere Streuung der Testergebnisse nach Live-Simulationen lässt annehmen, dass hierbei individuelles Wissen stärker zum Tragen kommt.

Die getroffenen Aussagen basieren auf den Auswertungen des vorhandenen Studienmaterials und sind lediglich als Vermutung zu werten. Die Testergebnisse resultierten aus in Inhalt und Schwierigkeit unterschiedlichen Patientenfällen und könnten auch durch die Zusammensetzung der Kohorten in den jeweiligen Simulationsarten hervorgerufen worden sein. Der Einfluss von personenspezifischen Charakteristika wie Alter, Bildungsniveau oder Selbsteinschätzungen wurde im Rahmen der Auswertung nicht untersucht. Zusammenhänge könnten somit auch aufgrund nicht in Betracht gezogener Drittvariablen hervorgerufen worden sein.

#### **4.2.2 Diskussion der Fragestellungen**

Die Performanzen in Konzeptuellem Wissenstest und Key Feature Test korrelierten stark positiv. Bei annähernd gleich großen Stichproben wurde das Ergebnis von Schmidmaier et al. (2013), schwach positive Korrelationen der Testergebnisse [28], nicht bestätigt. Jedoch war die Testkonzeption in beiden Studien zur Messung von strategischem und konzeptuellen Wissen unterschiedlich. Der hohe Koeffizient lässt ähnliche Strukturen innerhalb der Messinstrumente vermuten: beide Wissenstests beinhalteten Fragen für Diagnostik, Therapie und Symptomatik

zum Leitsymptom Atemnot. Inhaltlich spricht die stärkere Korrelation jedoch nicht unbedingt für einen kausalen Zusammenhang der beiden Wissensarten, sodass sich das vorliegende Ergebnis eher aus einer Strukturähnlichkeit der Messinstrumente erklären dürfte. Medizinisches Wissen sollte auch weiterhin hinsichtlich der in der Literatur beschriebenen Arten kategorisiert werden (s. 1.2.2), deren Kombination sich letztlich positiv in Diagnosekompetenzen zeigt.

Die Performanzen aus Wissenstests und Simulationen zu korrelieren diene der Überprüfung, ob anhand der Messung von klinischen Diagnosekompetenzen in simulierten Anamnesegesprächen gleichsam Aussagen hinsichtlich der Ausprägung verschiedener Arten medizinischen Wissens getroffen werden können. Die Testergebnisse aus Wissenstests und Simulationen korrelierten moderat positiv. Gemäß Jones et al. (2019) zeigt sich die Anwendung von Wissen in Kompetenz [18]. Dies konnte anhand der Testergebnisse bestätigt werden. Die stärkere Korrelation zwischen Simulationen und konzeptuellem Wissen als strategischem Wissen ist inhaltlich begründbar. Aus den Simulationen wurden Verdachtsdiagnosen und Begründungen dieser ausgewertet. Studierende wurden gebeten, konzeptuelles Wissen, Fakten, anzuwenden. Wissen, als Basis klinischer Diagnosekompetenzen, kann somit anhand von Wissenstests oder Simulationen gleichsam erfasst werden. Eine standardisierte, qualitative Auswertung der aus den Simulationen erfassten Antworten für Untersuchungsmaßnahmen und Therapie war aufgrund des Antwortformats nicht möglich (s. 4.3). Interessant wäre, ob unter Einbeziehung dieser, stärkere Korrelationen mit den Testergebnissen des Key Feature Tests, und somit der Anwendung von strategischem Wissen, bestanden hätten (s. 4.4). Die Betrachtung der Testergebnisse hinsichtlich Gruppe 1 und Gruppe 2, diene der Fragestellung, anhand welcher Simulationsart Diagnosekompetenzen besser erfasst werden können. Dies wurde aus den erhobenen Daten nicht geklärt. Die Korrelationen waren fallspezifisch und nicht aufgrund einer Simulationsart gegeben. Der sehr geringe Zusammenhang zwischen den Performanzen in Live- und Video-Simulation innerhalb einer Gruppe weist erneut auf den Einfluss der Jensen et al. (2019) beschriebene Case Specificity für Diagnosekompetenzen hin: Komplexität und Inhalt eines Patientenfalls scheinen Diagnosekompetenzen zu beeinflussen und qualifizieren diese als nicht generalisierbar [21]. Gemäß der Auswertung der Testergebnisse wäre bei Studierenden im Praktischen Jahr die Performanz in Simulationen weitestgehend unabhängig von konzeptuellem und strategischem Wissen (s. Tabellen 18-20). Dass jedoch mit zunehmender Expertise bessere Performanz in den Wissenstests nicht mit besserer Performanz in den Simulationen assoziiert ist, ist unwahrscheinlich. Eher könnte medizinisches Wissen aufgrund der von Schmidt und Rikers (2007)

beschriebenen Enkapsulierung bereits stärker verknüpft und deshalb nicht mehr klar differenziert messbar sein [38]. Kompetenzen wären dann mehr als globales Ergebnis der Expertise zu werten als hinsichtlich einzelner Komponenten zu klassifizieren.

Assoziationsfaktoren für Performanz in Simulationen zu überprüfen diene dem Gewinn zukünftiger Implikationen für die Ausbildung von Medizinstudierenden. Anhand der Testergebnisse scheint die Art der Simulation die Messung von Performanz zu beeinflussen. Die höheren Wahrscheinlichkeiten für korrekte Diagnosen nach Live-Simulationen (ausgenommen Lehner) könnten sich aus dem von Christensen et al. (2019) beschriebenen unmittelbaren Feedback [22] und einer gezielteren, individuelleren Kommunikation erklären. Die spezifischen Diagnosen zeigten sich sensibler in Bezug auf die Simulationsart: Unterschiede in den Wahrscheinlichkeiten für korrekte Diagnosen waren hierbei größer. Erneut war daher zu erkennen, dass die Komplexität eines Patientenfalls in Kombination mit der Simulationsart die Messung von klinischen Diagnosekompetenzen zu beeinflussen scheint.

Die Untersuchung eines Effekts der Reihenfolge diene dazu, ob das Vorschalten einer Simulationsart, insbesondere von Video-Simulationen, aufgrund der strukturierten computerbasierten Lernumgebung, wie von Kopp et al. (2006) beschrieben Hinweisreize (Cueing) generiert [78], die sich in der Messung von Performanz zeigen. Weder anhand statistischen Modells noch deskriptiv ergaben sich konsistente Assoziationen zwischen dem Vorschalten einer Simulationsart und der Performanz in Simulationen. Zusammenhänge waren fallspezifisch sowie bei Gruppierung der Patientenfälle erkennbar, aber nicht aufgrund der Simulationsart. Die Quantifizierung eines möglichen Effekts der Vorschaltung zeigte sehr breite Konfidenzintervalle und somit eine große Unsicherheit der Ergebnisse. Bei allen Patientenfällen führte das Vorschalten von Simulationen anhand der statistischen Modelle nicht zu eindeutig positiven oder negativen Effekten. Basierend auf den Daten und der hohen Unsicherheit können beide Szenarien somit nicht ausgeschlossen werden. Die beobachteten Assoziationen sind statistisch als nicht gesichert anzusehen und die Aussagen über mittlere Veränderungen somit mit Vorsicht zu werten. Die Ergebnisse nach Studiumsstatus aufzuteilen generiert aufgrund geringerer Fallzahlen noch größere Unsicherheiten, weshalb die Auswertung lediglich zur Vollständigkeit im Anhang beigefügt ist (s. Anhang 9).

Aus den Testergebnissen zeigte sich, dass klinische Diagnosekompetenzen durch den Aufbau und die Entwicklung medizinischer Expertise gefördert werden und als nicht generalisierbar, somit fallspezifisch, zu erachten sind [21]. Die bessere Diagnosequalität nach Anamnesegesprächen in Live-Simulation unterstützt die Empfehlungen von Bacher-Meck (2015), weiterhin

direkte Kommunikation und nicht computergestützte Anamnesesysteme in der Klinik anzuwenden [11]. Für die Messung von klinischen Diagnosekompetenzen bei Medizinstudierenden sind die breiten Möglichkeiten der Video-Simulationen zu nutzen. Anhand der Testergebnisse scheint das Format per se keine inhaltlichen Hinweisreize (Cueing) zu generieren und ist als sinnvolles, effektives und langfristig gesehen kosteneffizientes Messinstrument für die Ausbildung von Medizinstudierenden zu erachten. Lernunterstützende Maßnahmen können je nach Status medizinischer Expertise standardisiert integriert werden und somit durch Scaffolding wichtige Aspekte für die Ausbildung von Diagnosekompetenzen gefördert werden [67]. War das simulierte Anamnesegespräch im Rahmen der medizinischen Ausbildung bis dato vornehmlich Bestandteil eines OSCE [86], dessen Durchführung mit hohem organisatorischem Aufwand verbunden ist, suggerieren die Ergebnisse der Studie, dass auch ein isolierter Einsatz von Video-Simulationen nicht nur effizient sondern auch effektiv eingesetzt werden kann.

### **4.3 Diskussion des Studienmaterials und der Methoden**

Die Item- und Skalenanalyse der Wissenstests sollte den Autoren der Messinstrumente Aufschluss über das Testmaterial gewährleisten. In beiden Wissenstests waren vereinzelt Decken- und Bodeneffekte bei den Frageitems erkennbar. Diesbezüglich bedarf es vor allem im Key Feature Test einer Überarbeitung. Jedoch werden weiterhin Frageitems breit gestreuter Schwierigkeiten Daseinsberechtigung in beiden Wissenstests finden, wenn sie inhaltlich Entscheidendes abdecken und somit zur Testvalidität beitragen. Unter Berücksichtigung der Trennschärfe sind beide Wissenstests mit neuen Fragen zu ergänzen, um in den kommenden Studienphasen wieder ein Testmaterial aus 40 Fragen und somit auch vergleichbaren Umfangs einsetzen zu können. Dabei gilt es inhaltlich präziser darauf zu achten, ob diese Wissen bezüglich Fakten oder strategischem Vorgehen prüfen, um durch schärfere Testkonzeption eine geringere Korrelation der Testergebnisse zu erzeugen. In wieweit aussagekräftige, reliable Messungen anhand der einzelnen Skalen erfolgten, ist unklar. Die Skalen in den Wissenstests beinhalteten unterschiedlich viele und teilweise sehr wenige Items. Hinsichtlich des Antwortformats erwiesen sich sowohl Single Choice Typ A als auch Pick-N Multiple Choice als adäquat. Eine Anpassung auf eine gleiche Anzahl gegebener Antwortitems in beiden Wissenstests könnte darüber hinaus beide Messinstrumente vergleichbarer gestalten und die Güte der Aussagen verbessern.

Die deskriptive Auswertung der Patientenfälle zeigte, dass die Vorbefunde in den Protokollen adaptiert und teilweise ergänzt werden müssen. Diese ex ante nicht erkannten Ungenauigkeiten

fürten auch dazu, dass der Fall Wagner nachträglich für die Auswertungen ausgeschlossen werden musste. Die Anzahl der Patientenfälle war in den Gruppierungen somit nicht mehr gleich, was Ergebnisse verzerrt und damit die Aussagekraft der Auswertungen reduziert haben könnte. Die standardisierten Patienten erzeugten eine komplexe Übungssituation, da diese in Abgrenzung zum Peer-Learning in Anamnesegruppen und getreu der Klinik, den Teilnehmenden bei Erstkontakt nicht bekannt waren. Durch die vorgegebene Zeit von 8:30 Min. sollte dem Handlungsdruck in klinischem Umfeld nachgekommen werden sowie beide Simulationsarten vergleichbarer gestaltet werden. Aufbau und Umfang eines Anamnesegesprächs sind jedoch stark kontextspezifisch, selbst innerhalb eines Leitsymptoms vom Krankheitsbild, der Komplexität eines Patientenfalls und der Persönlichkeit des Diagnostizierenden abhängig und deshalb schwer über ein definiertes Zeitfester zu fixieren. Dies zeigte sich vor allem anhand der Live-Simulationen, das Gespräch wurde oft vor Ablauf der 8:30 Min. beendet. So ist zu evaluieren, in wieweit eine genau festgelegte Zeit in den kommenden Phasen weiter Teil des Studiendesigns bleibt.

Durch das Anamnesegespräch sollten Diagnosekompetenzen anhand von Diagnosequalität in Akkuratheit und Effizienz erfasst werden (s. Abbildung 2): Akkuratheit durch Angabe einer Verdachtsdiagnose und Begründung dieser, Effizienz anhand fallspezifischer Untersuchungen und Therapiemaßnahmen unter Berücksichtigung dafür erforderlicher Kosten. Die Verdachtsdiagnose wurde aus einer Liste von 239 Items über ein Drop-Down Menü ausgewählt. Durch diese Liste sollten in Anlehnung an Long-Menu Antwortformate Hinweisreize (Cueing) zur Auswahlunterstützung ausgeschlossen werden [92]. Die Liste wurde neu konzipiert und bedarf inhaltlich einer Schärfung hinsichtlich gewünschter korrekter Diagnosen, um in den kommenden Studienphasen die Auswertbarkeit der einzelnen Patientenfälle qualitativ zu verbessern. Die Angabe von Begründung, weiteren Untersuchungen sowie Therapiemaßnahmen erfolgte in Freitexteingabe. Jedoch erzeugte das Antwortformat erhebliche Qualitätsverluste für die Messbarkeit klinischer Diagnosekompetenzen. Inhalt und Umfang gegebener Antworten waren derart variabel, dass die tatsächliche Qualität der Aussagen nicht standardisiert auszuwerten war. Deshalb wurden auch letztlich die Parameter der Effizienz, Untersuchungen und Therapiemaßnahmen, nicht ausgewertet. Angaben hierfür sollten fallspezifisch, bedarfsgerecht, unter Berücksichtigung des finanziellen Aspekts der Kosteneffizienz erfolgen. Die gegebenen Antworten glichen mehr einer Reproduktion konzeptuellen Wissens hinsichtlich der diagnostizierten Krankheitsbilder als einem in dem spezifischen Patientenfall richtigen, strategischen Vorgehen. Zu Recht stellt sich hierbei die Frage, ob dies eventuell auch als Kennzeichen fehlender Expertise zu werten ist. Wenden Studierende im Klinischen Studienabschnitt mehr konzeptuelles Wissen an und zeigt

sich bei Studierenden im Praktischen Jahr bereits fallspezifischer mehr strategisches Vorgehen? Eine Überprüfung der gegebenen Antworten diesbezüglich konnte keine auswertbaren Tendenzen erkennen lassen, vermutlich auch dem Antwortformat geschuldet.

Effizienz in medizinischem Handeln in Bezug auf anfallende Kosten im weiteren Behandlungsverlauf schien anhand der gegebenen Antworten nicht berücksichtigt worden zu sein. Einerseits wurden den Teilnehmenden diesbezüglich zwar keine Auskünfte bereitgestellt, andererseits zeigte sich daraus die Erkenntnis, dass Medizinstudierenden die Kosten für Untersuchungen und Therapien weitestgehend unbekannt sind. Medizinstudierende werden über die finanzielle Relevanz, wie Kosten-Nutzwert ärztlicher Entscheidungen, unzureichend unterrichtet. Nicht mangelndes pathophysiologisches Wissen oder mangelnde Fähigkeiten der Studierenden Kausalzusammenhänge im Rahmen eines Diagnoseprozesses zu erkennen, sondern die Konzeption des Antwortformats war auch für das schlechte Abschneiden bei den Begründungen größtenteils mitverantwortlich. Durch die Freitexteingabe flossen auch intrinsische Faktoren wie Teilnahmemotivation bei der Bearbeitung mit ein und könnten somit die tatsächlichen Diagnosekompetenzen verzerrt dargestellt haben. Deshalb gilt es für die kommenden Studienphasen die Antwortformate für Begründung, Untersuchungen und Therapiemaßnahmen zu ändern. Anstelle der Freitexteingabe könnte hierfür eine Konzeption gemäß Pick-N Multiple Choice in Anlehnung an das Studiendesgin im Teilprojekt 6 der Cosima Forschergruppe verwendet werden [98]. Die verbesserte Kodierungsqualität durch Standardisierung könnte dann klinische Diagnosekompetenzen bei Medizinstudierenden genauer abbilden.

Anamnesegespräche in einer simulierten Lernumgebung und nicht direkt an echten Patienten zu trainieren sowie die Konzeption der Simulationen wurden von den Medizinstudierenden durchwegs positiv bewertet. Einziger Kritikpunkt war die fehlende Auskunft über die richtigen Lösungen der einzelnen Patientenfälle. Damit Fehler produktiv sind und Diagnosekompetenzen dadurch verbessert werden können, bedarf es jedoch einer kritischen Evaluation und Diskussion der Ergebnisse. Diese sollte möglichst zeitnah erfolgen, damit der Grund einer falschen Antwort nachvollzogen und fehlerhafte Wissensnetzwerke gleichzeitig korrigiert werden können [99], [5].

Für die Beantwortung der Fragestellungen 1 und 2 wurden lediglich Korrelationen durchgeführt. Der Einfluss von personenspezifischen Charakteristika wie Selbsteinschätzung der Studierenden wurde außer Acht gelassen. Als einzige Einflussvariable wurde der Studiumsstatus untersucht. Unter dieser Kategorisierung ergaben sich zwei ungleich große Kohorten von 47 Studierenden im Klinischen Studienabschnitte und 39 Studierenden im Praktischen Jahr. Die

dritte Fragestellung hätte teilweise auch anhand der deskriptiven Auswertung der Patientenfälle beantwortet werden können. Fokus hierbei war jedoch nicht das Studienmaterial, sondern mögliche Transfereffekte aus simulierten Lernumgebungen zu ermitteln, weshalb umfangreichere statistische Auswertungen durchgeführt wurden.

Durch das 2 x 2 Design der Simulationsphase (s. 2.1) wurde unter Aufteilung nach Studiumsstatus der eigentlich große Stichprobenumfang von  $N = 86$  relativiert. Jede der 4 Zellen beinhaltete insgesamt weniger als 30 Personen, was die Generalisierbarkeit der Aussagen einschränkt. Hauptursache hierfür lag vor allem in der Teilnehmerfindung. Studierende im Praktischen Jahr mussten für die Teilnahme eine Befreiung vom Stationsdienst beantragen. Für die zukünftigen Studienphasen wird dies durch ein offizielles Freistellungsgesuch des DAM erleichtert und das Projekt konnte durch die Phase 1 bereits in einigen Kliniken in den Stationsablauf integriert werden. Dies gewährleistet auch eine größere Effizienz für zukünftige Erhebungen.

Die Kohorte der Studierenden im Klinischen Studienabschnitt war sehr breit über 3 Studienjahre gefasst. Hierdurch konnte eine gute Repräsentativität der Stichprobe erreicht werden. Jedoch mussten innerhalb dieser Kohorte die Diagnosekompetenzen bereits unterschiedlich stark ausgeprägt gewesen sein. Für allgemeine Messungen von klinischen Diagnosekompetenzen bei Medizinstudierenden ist ein breit gefasstes Kollektiv aufgrund der Repräsentativität durchwegs zur befürworten. Für die Evaluation von Messinstrumenten in Bezug auf Qualität und Messgenauigkeit hätte eine einheitlichere Kohorte vermutlich eine bessere Aussagekraft gehabt. Der Informationsbias ist als gering einzuschätzen, da aufgrund des Titels der Studie, „Üben von Anamnesegesprächen in Simulationen“, keinerlei inhaltlichen Auskünfte gegeben wurden, wodurch Vorinformationen zum Leitsymptom Atemnot hätten eingeholt werden können. Aus Effizienzgründen gilt es die gegenwärtige Ausschlussquote von über 5 % der erfassten Testergebnisse in den kommenden Studienphasen zu reduzieren.

#### **4.4 Implikationen und zukünftige Forschungen**

Obgleich die Ergebnisse aus der Evaluationsphase neu konzipierter Messinstrumente stammen, sind bereits umfangreiche Implikationen für die Ausbildung von Medizinstudierenden sowie für zukünftige Forschungen zu erkennen. Der wissenschaftliche Mehrwert dieser Arbeit liegt vor allem in den Messungen aus den Simulationen. Zentrale Erkenntnis daraus ist, dass für die Ausbildung von Medizinstudierenden der Einsatz von Video-Simulationen zukunftssträchtige



Chancen für das Trainieren von Anamnesegesprächen bietet. Zeit- und ortsunabhängig ermöglichen sie den Aufbau und die Messung von klinischen Diagnosekompetenzen bei Medizinstudierenden. Je nach Status medizinischer Expertise können individuell, kosteneffizient und mit geringem Aufwand, lernunterstützende Maßnahmen in Form von Scaffolding integriert und der Umfang an Cognitive Load auf Medizinstudierende moduliert werden [67]. Darüber hinaus bieten Video-Simulationen die Möglichkeit durch Implementierung eines Lösungstools kostengünstig und standardisiert qualitativ hochwertiges ärztliches Feedback an Medizinstudierende weiter zu geben. Lernerfolg durch den Einsatz von Lösungsbeispielen wurde bereits von Kopp et al. (2007) aufgezeigt [19], kann effektiv zur Entwicklung von Diagnosekompetenzen beitragen und wird entscheidend durch die Qualität der Supervision beeinflusst [100]. Des Weiteren sollte aufgrund der Testergebnisse evaluiert werden, in wieweit gesundheitsökonomische Aspekte in die Ausbildung von Medizinstudierenden etabliert werden sollten. Denn der Erwerb breit ausgelegter medizinischer Expertise ist entscheidend für bedarfsorientiertes, kosteneffizientes Handeln sowie Patientenoutcome. Um aus den neu konzipierten Instrumenten letztlich valide Aussagen zur Messung von klinischen Diagnosekompetenzen bei Medizinstudierenden treffen zu können, bedürfen diese einer kontinuierlichen Evaluation sowie mehrfachen Testungen im Rahmen der kommenden Studienphasen.

Anhand von zukünftig vollständigen Testergebnisse aus den Simulationen könnte in einem nächsten Schritt geprüft werden, in welchem Ausmaß strategisches und konzeptuelles Wissen mehr mit Akkuratheit und Effizienz von Diagnosen zusammenhängen. Vermutlich bestehen, gemäß dem in der vorliegenden Arbeit verwendeten Modell für Diagnosekompetenzen (s. 1.3.5) stärkere Zusammenhänge zwischen konzeptuellem Wissen und Diagnoseakkuratheit, sowie strategischem Wissen und Diagnoseeffizienz. Die Anwendung von konzeptuellem und strategischem Wissen könnte auch für die Überprüfung von kognitiver Flexibilität als Zeichen von Expertise herangezogen werden. Hierbei wäre zu untersuchen, ob Studierende bei Bearbeitung der Patientenfälle in der Lage sind, fallspezifisch strategisch zu handeln, oder lediglich konzeptuelles Faktenwissen über Krankheitsbilder pauschal anwenden. Fallspezifisches Handeln müsste dann mehr bei Studierenden im Praktischen Jahr als bei Studierenden im Klinischen Studienabschnitt erkennbar sein. Inwieweit das Praktische Jahr per se maßgeblich zur Entwicklung von Diagnosekompetenzen beiträgt, könnte anhand einer Kohorte aus Teilnehmenden im ersten und dritten Trimester des Praktischen Jahres untersucht werden.

Die Aufteilung nach Studiumsstatus bietet insgesamt ein breites Feld an Möglichkeiten, um anhand der in der vorliegenden Arbeit untersuchten Messinstrumente verschiedene Theorien hinsichtlich der Entwicklung von Expertise zu untersuchen.

In einem nächsten Schritt sind die gefundenen Zusammenhänge statistisch zu replizieren und damit für die Ausbildung von Medizinstudierenden zu objektivieren. Die Empfehlung ist, den Fokus auf die Video-Simulationen zu intensivieren. Aufwand und Kosten der Konzeption haben das Potenzial, sich in einem doppelt positiven Nutzen widerzuspiegeln. Video-Simulationen sind langfristig kosteneffizient und dienen dadurch der Kosteneinsparung. Video-Simulationen können in der Lehre eingesetzt werden Anamnesegespräche effektiv zu trainieren, dadurch Diagnosekompetenzen fördern und somit Fehldiagnosen reduzieren. Weniger Fehldiagnosen verbessern Behandlungsergebnisse und Patientenzufriedenheit. Weniger inadäquate Untersuchungen und zusätzlich erforderliche Therapiemaßnahmen generieren darüber hinaus Kosteneinsparungen im Gesundheitssystem. Dies rechtfertigt jegliche zukünftige wissenschaftliche Investition in die Weiterentwicklung simulierter Lernumgebungen.

„Errare humanum est“ [2]. Wo durch Konzeption Neues entsteht, Menschen handeln, werden Fehler begangen. „Sed in errare perseverare diabolicum“ [2]. Lediglich auf Fehlern zu beharren, nicht zu handeln, Innovationen zu unterbinden wäre infaust. Dies gilt nicht nur für die Konzeption von Messinstrumenten für eine bessere Ausbildung von Medizinstudierenden, sondern auch im weiteren klinischen Alltag für den Umgang mit Fehldiagnosen.

## 5 Anhang

Anhang 1:	Forschungsprogramm Cosima Teilprojekt 4 .....	69
Anhang 2:	Beispielfragen aus Konzeptuellem Wissenstest.....	69
Anhang 3:	Key Feature Fall aus Key Feature Test.....	70
Anhang 4:	Fallvignette aus Simulationen.....	72
Anhang 5:	Musterlösung Patientenfälle.....	77
Anhang 6:	Anamnesebogen .....	80
Anhang 7:	Überprüfung der Messinstrumente nach Studiumsstatus.....	81
Anhang 8:	Auswertungen der Forschungsfrage 3b) nach Studiumsstatus .....	84

## Anhang 1: Forschungsprogramm Cosima Teilprojekt 4

Phase	Ziel
Studie 1	Validierung und Entwicklung der Messinstrumente und des Studienmaterials
Studie 2a	Prüfung der Effekte von übernommenen Rollen (Arzt/Patient)
Studie 2b	Prüfung der Effekte von Scaffolding durch Reflexionsphasen
Studie 3	Prüfung der Effekte von übernommenen Rollen in Kombination mit Scaffolding

## Anhang 2: Beispielfragen aus Konzeptuellem Wissenstest

### Frage 14:

Welcher Laborparameter ist zur Diagnose einer Herzinsuffizienz am besten geeignet?

(Bitte kreuzen Sie eine Antwort an!)

- (A) Troponin T
- (B) D-Dimere
- (C) Myoglobin
- (D) Hochsensitives C-Reaktives Protein (CRP)
- (E) X B-Typ natriuretisches Peptid (BNP)

### Frage 25:

Welche Maßnahmen führen bei COPD zu einer Lebensverlängerung?

(Bitte entscheiden Sie bei jeder Aussage, ob diese zutrifft oder nicht)

- (A) ☐ Ja ☒ Nein Inhalative Kortikosteroide
- (B) ☐ Ja ☒ Nein Langwirksame Beta-2-Sympathikomimetika
- (C) ☒ Ja ☐ Nein Nikotinkarenz
- (D) ☒ Ja ☐ Nein Sauerstofflangzeittherapie
- (E) ☐ Ja ☒ Nein Systemische Kortikosteroide

### Anhang 3: Key Feature Fall aus Key Feature Test

#### Fall 4 (Pneumonie)

---

Es ist Donnerstag Vormittag in der Hausarztpraxis, wo Sie als Ärztin/Arzt in Weiterbildung tätig sind. Zu Ihnen kommt die 48-jährige Sarah Brecht. Sie berichtet Ihnen von starken rechtsseitigen Thoraxschmerzen seit heute morgen. Diese hätten über Nacht langsam zugenommen und würden inzwischen das Atmen deutlich erschweren. Insgesamt bekomme sie schon seit gestern Nachmittag nicht mehr so richtig Luft, fühle sich überhaupt sehr krank und habe hohes Fieber (bis 40,3°C).

**Frage 13:** Was ist Ihre wahrscheinlichste Verdachtsdiagnose?

- a. Anaphylaktische Reaktion
- b. Bronchial-Ca
- c. Bronchitis
- d. Cholezystitis
- e. Grippaler Infekt
- f. Hinterwandinfarkt
- g. Infektexazerbierte COPD
- h. Pankreatitis
- i. Pathologische Fraktur
- j. Pneumonie**
- k. Spondylodiszitis
- l. Typhus

---

**Frage 14:** Welches weitere Symptom ist am wichtigsten zu prüfen?

- a. Hypersonorer Klopfschall
- b. Schwindel
- c. Kopfschmerzen
- d. Schlafstörungen
- e. Hyperventilation**
- f. Nykturie
- g. Gelenkschmerzen
- h. Übelkeit
- i. Leistungsknick
- j. Sprachveränderungen
- k. B-Symptomatik
- l. Inspiratorischer Stridor

---

**Frage 15:** Was ist in dieser Situation die wichtigste diagnostische Maßnahme zur Planung des weiteren Procederes?

- a. Blutkulturen
- b. Blutdruckmessung**
- c. Blutgasanalyse
- d. CRP-Schnelltest
- e. EKG
- f. Elektrolyte im Serum
- g. Gastroskopie
- h. Kleines Blutbild**
- i. Bodyplethysmographie
- j. D-Dimere bestimmen
- k. Röntgen-Thorax in 2 Ebenen
- l. Abdomensonographie

---

**Frage 16:** Sie erheben bei Frau Brecht die folgenden pathologischen Befunde: CRP 73 mg/l (Referenzbereich: < 8 mg/l), Rasselgeräusche links basal, RR 110/75 mmHg, Puls 102/Min. regelmäßig, AF 25/Min., Temperatur 39,2°C, SpO<sub>2</sub> 99 % bei Raumluft.

Was ist in dieser Situation die wichtigste therapeutische Maßnahme?

- a. Betablocker (Bisoprolol)
- b. Abwartendes Offenlassen
- c. Analgetika (Ibuprofen)
- d. Klinikeinweisung
- e. Atemtherapie
- f. Glukokortikoide (Prednisolon) i.v.
- g. Namentliche Meldung an das Gesundheitsamt
- h. Antibiotika (Amoxicillin + Clavulansäure)**
- i. Schleimlöser
- j. Überweisung an Infektiologie
- k. Antikoagulation (Heparin i.m.)
- l. Überweisung an Radiologie

## Anhang 4: Fallvignette aus Simulationen

Sie sind Anette Klein, eine etwa 55-jährige Frau, studierte Biologielehrerin, arbeiten aber seit der Geburt der Kinder nicht mehr in Ihrem Beruf, sondern helfen bei Ihrem Mann in seiner Anwaltskanzlei aus. Sie haben zwei erwachsene Töchter, hatten zwischen den beiden Schwangerschaften eine Fehlgeburt. Die Ausschabung nach dieser Fehlgeburt ist die einzige Operation, die Sie jemals hatten. Gesundheitlich sind Sie seit einiger Zeit heftig mit den Wechseljahren beschäftigt, die Ihnen starke Beschwerden machen. Vor allem die Schweißausbrüche finden Sie schlimm, die Sie für die Infekte verantwortlich machen, die Sie seit einiger Zeit immer wieder haben. Nachts ist es zwar nicht so schlimm wie tagsüber, aber trotzdem schwitzen Sie so etwa einmal die Woche nachts so stark, dass Sie sich umziehen müssen. Außerdem haben Sie eine Anämie (Blutarmut), weil Sie immer noch sehr starke Menstruationsblutungen haben. Ansonsten sind Sie gesundheitlich eigentlich gut beieinander - seit etwa 10 Jahren nehmen Sie Tabletten gegen Bluthochdruck, damit kommen Sie gut zurecht. Außer den Blutdrucktabletten nehmen Sie noch ein paar naturheilkundliche Präparate gegen die Wechseljahresbeschwerden und vielleicht so ungefähr einmal im Monat etwas gegen Kopfschmerzen. Sie haben im letzten Jahr so ungefähr 5 kg zugenommen, was Sie aber nicht weiter wundert - Sie essen einfach gerne zurzeit.

Sie rauchen nicht, trinken Alkohol nur in Maßen und machen so ungefähr zwei Mal pro Woche Sport - meistens Zumba oder so etwas. Ihre Eltern leben noch - sie sind zwar alt, aber noch ganz rüstig, beide haben hohen Blutdruck, der Vater ist zuckerkrank und die Mutter hat Rheuma. Stark beeinflusst hat Sie der Tod Ihrer Tante an Brustkrebs, als Sie ein Teenager waren. Deshalb gehen Sie auch regelmäßig zur Vorsorge bei Ihrer Hausärztin, messen Ihren Blutdruck (der ist immer in Ordnung, so um 120/80 rum), wissen, dass Ihre Schilddrüse in Ordnung ist und sind sehr gewissenhaft, was Ihre Gesundheit angeht.

Gestern war ein ganz komischer Tag. In der Früh haben Sie sich noch ganz gut gefühlt, waren sogar noch beim Joggen, es war alles in Ordnung. Am Nachmittag sind Sie dann innerhalb weniger Stunden furchtbar krank geworden, hatten Schüttelfrost, haben sich sehr krank und schwach gefühlt und haben gar nicht mehr so richtig Luft bekommen. In der Nacht haben Sie fürchterlich geschwitzt und hatten ständig so Alpträume, dass Sie ertrinken müssen.

In der Früh haben Sie dann den Notarzt gerufen, weil dann auch noch so Schmerzen beim Atmen dazu gekommen sind, Sie können gar nicht mehr richtig Luft holen.

### 55-jährige Patientin, fiebrig, kurzatmig, blass

**„Guten Morgen Frau Klein. Ich bin heute für Sie in der Notaufnahme zuständig. Was kann ich denn für Sie tun?“**

"Mir geht es so schlecht seit gestern Nachmittag. Gestern früh habe ich mich noch fit gefühlt, ich war sogar noch beim Joggen, aber am Nachmittag habe ich dann innerhalb ein paar Stunden so hohes Fieber und ganz schreckliche Kopfschmerzen bekommen. Und als ich heute Nacht im Bett lag, hatte ich das Gefühl, dass ich gar nicht richtig Luft kriege, ich hatte immer so Alpträume, dass ich ertrinke."

**Haben Sie die Beschwerden zum ersten Mal?**

Ja, sowas habe ich überhaupt noch nie erlebt.

**Haben Sie auch Schmerzen?**

Ja, beim Atmen tut es weh, ganz scheußlich. Wenn ich die Luft anhalte, ist es besser, aber das geht ja nicht die ganze Zeit.

<b>Sind bei Ihnen irgendwelche Vorerkrankungen bekannt?</b>	Ich habe ziemlich starke Wechseljahresbeschwerden immer noch. Also so Hitzewallungen, Schweißausbrüche, Schlafstörungen. Und auch noch ziemlich starke Blutungen. Deswegen habe ich auch eine Anämie. Und vor einem Jahr ist ein Bluthochdruck festgestellt worden, da nehme ich jetzt Tabletten.
<b>Haben Sie häufig Infekte, gegen die Sie auch Antibiotika nehmen müssen?</b>	Naja, also ich bin schon in letzter Zeit manchmal krank, ich denke, das kommt von den Wechseljahren, weil ich da oft so Schweißausbrüche habe und dann wird es halt manchmal auch kalt in den nassen Klamotten. Aber ein Antibiotikum habe ich nur 1 Mal gebraucht in den letzten zwei Jahren.
<b>Wurden Sie schon mal operiert?</b>	Ich hatte vor 20 Jahren mal eine Ausschabung nach einer Fehlgeburt. Sonst nichts.
<b>Sind Ihre Eltern oder sonst jemand in Ihrer näheren Verwandtschaft sehr jung gestorben?</b>	Nein, meine Eltern leben noch und sind zwar inzwischen alt, aber eigentlich noch ganz rüstig. Sonst weiß ich auch von niemand, der irgendwie auffallend jung gestorben wäre.
<b>Ist in Ihrer Familie jemand am plötzlichen Herztod gestorben?</b>	Nein, davon weiß ich nichts.
<b>Wie fühlen Sie sich jetzt?</b>	Schlecht fühle ich mich. Ich kriege einfach nicht genug Luft, und diese Schmerzen sind auch wirklich unangenehm. Und ganz schwach und krank fühle ich mich.
<b>Hat sich Ihr Gewicht verändert in den letzten Wochen?</b>	Ich habe in diesem Jahr leider schon so um die 5 kg zugenommen. Aber das liegt an dem, was ich esse, ist jetzt kein großes medizinisches Rätsel ehrlich gesagt.
<b>Leiden Sie unter Nachtschweiß?</b>	Heute Nacht habe ich sehr stark geschwitzt. Ansonsten auch manchmal, das sind halt diese Hitzewallungen. Zum Glück ist das nachts nicht so ein riesengroßes Problem. Vielleicht einmal in der Woche muss ich mich nachts umziehen.
<b>Nehmen Sie regelmäßig Medikamente? Oder nehmen Sie vielleicht Bedarfsmedikamente? Also so etwas wie Allergietabletten oder Schmerzmittel?</b>	Meine Blutdruckmedikamente. Und ein paar naturheilkundliche Präparate gegen die Wechseljahresbeschwerden. Manchmal eine Kopfschmerztablette, aber normalerweise höchstens einmal im Monat. Gestern und heute habe ich aber schon was genommen, weil ich so schreckliche Kopfschmerzen hatte. Auch heute Nacht. Danach habe ich dann fürchterlich geschwitzt.
<b>Haben Sie heute ausreichend gegessen und getrunken?</b>	Heute nur ganz wenig gegessen und ein bisschen was getrunken. Ich habe gar keinen Appetit.



<b>Rauchen Sie oder haben Sie früher geraucht?</b>	Nein, zum Glück habe ich das nie angefangen.
<b>Wieviel Alkohol trinken Sie ungefähr?</b>	Schon seit ein paar Jahren trinke ich so gut wie gar keinen Alkohol mehr. Ich habe einfach gemerkt, dass mir das nicht gut tut. Also höchstens mal ein Glas Sekt alle paar Monate oder so.
<b>Haben Sie gestern viel Alkohol getrunken?</b>	Nein, gestern nichts.
<b>Haben Sie einen Beruf erlernt? Wenn ja, welchen?</b>	Ich habe Lehramt Biologie studiert.
<b>Sind Sie berufstätig?</b>	Ja, ich arbeite bei meinem Mann in der Kanzlei mit.
<b>Haben Sie bemerkt, dass Sie weniger leistungsfähig sind als früher?</b>	Also seit gestern bin ich überhaupt nicht mehr leistungsfähig. Ich komme ja kaum die Treppe rauf. Ansonsten bin ich jetzt im Wechsel auch ganz grundsätzlich nicht mehr so leistungsfähig wie früher - weder körperlich noch seelisch. Aber ich hoffe, das wird wieder besser, wenn ich endlich mal durch bin.
<b>Sind Sie verheiratet?</b>	Ja, seit 30 Jahren schon.
<b>Haben Sie Kinder?</b>	Ja, zwei Töchter, die sind aber schon erwachsen.
<b>Schlafen Sie gut?</b>	Normalerweise schlafe ich ganz gut. Heute Nacht war eine Katastrophe.
<b>Haben Sie Probleme beim Stuhlgang oder beim Wasserlassen?</b>	Nein, das ist alles ok.
<b>Haben Sie die Beschwerden nur bei Belastung, oder auch in Ruhe?</b>	Jetzt momentan wird es bei Belastung deutlich schlechter, eigentlich kann ich mich überhaupt nicht belasten. Aber auch in Ruhe geht es mir nicht gut.
<b>Sind noch andere Symptome aufgetreten? Also zum Beispiel Herzrasen, Schwindel oder so?</b>	Auch bei der geringsten Belastung bekomme ich Herzrasen und Schwindel, Schweißausbrüche und mir wird schwarz vor den Augen.
<b>Hatten Sie in letzter Zeit Teerstuhl oder haben Sie Blut erbrochen?</b>	Nein, das hatte ich noch nie.
<b>Leiden Ihre Eltern oder Geschwister an chronischen Erkrankungen (z.B. Bluthochdruck, Diabetes etc.)</b>	Meine Eltern haben beide Bluthochdruck, mein Vater ist auch zuckerkrank und meine Mutter hat Rheuma.
<b>Haben Sie in letzter Zeit eine längere Flug-, Bus- oder Autoreise unternommen?</b>	Nein, ich war daheim.
<b>Sind Sie in den letzten Wochen operiert worden?</b>	Nein, ich bin noch nie operiert worden.
<b>Waren Sie in Ihrer Beweglichkeit eingeschränkt, z.B. durch einen Gips oder durch eine Krankheit mit Bettlägerigkeit?</b>	Nein, es war eigentlich alles in Ordnung.

<b>Sind Sie schwanger oder haben Sie vor kurzem ein Kind bekommen?</b>	Wollen Sie Witze machen? Ich bin 55! Ich habe 2 erwachsene Töchter, zwischen den Schwangerschaften hatte ich eine Fehlgeburt.
<b>Ist bei Ihnen eine Gerinnungsstörung bekannt?</b>	Nicht, dass ich wüsste.
<b>Hatten Sie schon mal eine Thrombose?</b>	Nein.
<b>Haben Sie Kaugummi gekaut, als es passiert ist?</b>	Nein.
<b>Haben Sie gemerkt, dass die Augen jucken oder die Nase läuft?</b>	Nein, da ist mir nichts aufgefallen.
<b>Sind bei Ihnen Asthma oder Allergien bekannt?</b>	Nein, nichts in der Richtung.
<b>Hatten Sie in den letzten Tagen mal Fieber oder Schüttelfrost?</b>	Wenn Sie mich jetzt so fragen, kann das gut sein, dass ich Fieber hatte. Ich hatte letzte Nacht auf jeden Fall Schüttelfrost und habe auch ziemlich geschwitzt.
<b>Haben Sie zur Zeit auch einen Lippenherpes?</b>	Es kann schon sein, dass da gerade einer kommt, es kribbelt so an der Oberlippe.
<b>Ist bei Ihnen hoher Blutdruck bekannt?</b>	Ja, seit zehn Jahren schon.
<b>Machen Sie regelmäßig Sport?</b>	Ich versuche, mindestens zweimal pro Woche zum Sport zu gehen. Meistens Zumba oder so was.
<b>Hatten Sie schon mal Beschwerden mit dem Herzen?</b>	Nein, noch nie.
<b>Haben Sie Schmerzen im Arm oder im Kieferbereich?</b>	Nein, da tut nichts weh.
<b>Haben Sie ein Kribbeln oder Ziehen in den Händen?</b>	Nein, mit den Händen ist alles in Ordnung.
<b>Sind Sie schon mal wegen psychischer Probleme in Behandlung gewesen?</b>	Vor einigen Jahren hatte ich mal das Gefühl, dass mir zu Hause die Decke auf den Kopf fällt. Ich bin dann eine Weile zu einer Gesprächstherapie gegangen, das hat mir ganz gut getan.
<b>Haben Sie in letzter Zeit nicht mehr so häufig das Haus verlassen, weil Sie Angst hatten, dass so etwas passieren könnte?</b>	Nein, wirklich überhaupt nicht.
<b>Sind die Beine dicker geworden?</b>	Das ist mir nicht aufgefallen.
<b>Hatten Sie als Kind mal eine Herzmuskelentzündung?</b>	Nein, davon weiß ich nichts.
<b>Ist bei Ihnen eine Muskelerkrankung bekannt?</b>	Nein, davon weiß ich nichts.
<b>Ist die Schilddrüse mal untersucht worden auf eine Über- oder Unterfunktion?</b>	Ja, immer mal wieder. Da ist aber alles in Ordnung.

<b>Haben Sie in letzter Zeit Drogen oder irgendwelche Fitnessbooster oder so etwas genommen?</b>	Nein, weder in letzter Zeit noch irgendwann sonst.
<b>Ist das ganz plötzlich gekommen oder hatten Sie in den letzten Tagen oder Wochen schon mal Atemnot?</b>	Das ist ganz plötzlich gekommen. Also so über 3-4 Stunden ungefähr. Davor habe ich mich noch ganz wohl gefühlt und jetzt fühle ich mich wirklich schwer krank.
<b>Müssen Sie husten?</b>	Ein bisschen, aber nicht schlimm. Der Husten ist auch ganz trocken.
<b>Können Sie flach liegen?</b>	Ja, das ist kein Problem.
<b>Haben Sie in den letzten Monaten gelegentlich mal Probleme mit Schwindel gehabt? Oder das Bewusstsein verloren?</b>	Schwindel ab und zu. Aber bewusstlos war ich nie.
<b>Hatten Sie ein Gefühl von Todesangst?</b>	Nein. Todesangst nicht. Aber scheußlich ist es.
<b>Haben Sie in letzter Zeit manchmal ein Gefühl von Herzrasen oder -stolpern?</b>	Herzrasen schon, das sind diese Wechseljahre. Aber da habe ich mir nichts dabei gedacht.
<b>Haben Sie in den letzten Wochen einen akuten Infekt gehabt? Husten oder Schnupfen oder eine Grippe oder so?</b>	Ein bisschen Husten hatte ich in den letzten Tagen. Nix schlimmes.
<b>Haben Sie in letzter Zeit sehr viel Stress gehabt?</b>	Nicht mehr als sonst eigentlich.
<b>Gehen Sie regelmäßig zur Vorsorge, also dem Gesundheitscheck beim Hausarzt?</b>	Ja, normalerweise schon. Da war ich auch erst vor 6 Wochen und es war nichts Außergewöhnliches festzustellen.
<b>Wissen Sie, wie hoch normalerweise Ihr Blutdruck ist?</b>	So 120/80 rum normalerweise
<b>Wissen Sie, ob irgendwelche Blutwerte schon einmal schlecht gewesen sind? Cholesterin oder so etwas?</b>	Keine Ahnung, meine Hausärztin hat nie etwas gesagt.
<b>Wie sieht denn der Urin aus? Also ist Ihnen aufgefallen, dass der eine komische Farbe hat? Besonders hell oder besonders dunkel oder braun oder rot oder so?</b>	Nein, da ist alles ganz normal
<b>Können Sie mir sagen, wie viel Sie ungefähr trinken am Tag?</b>	Relativ viel, bestimmt so 3-4 Liter am Tag.
<b>Haben Sie Geschwister?</b>	Ja, eine große Schwester.
<b>Haben Sie schon mal eine bösartige Erkrankung gehabt, also Krebs oder einen Tumor oder so etwas?</b>	Nein, zum Glück nicht. Ich gehe auch immer ganz regelmäßig zu den Vorsorgen. Meine Tante ist an Brustkrebs gestorben, als ich ein Teenager war. Das hat mich stark beeinflusst damals.
<b>Sind bei Ihnen in der Familie irgendwelche Erbkrankheiten bekannt?</b>	Davon weiß ich nichts.

<b>Hatten Sie schon einmal einen Pneumothorax oder sind Sie schon einmal an der Lunge operiert worden?</b>	Nein, zum Glück noch nie!
<b>Hatten Sie schon einmal einen Schlaganfall?</b>	Also so alt bin ich doch noch gar nicht. Nein, natürlich nicht.
<b>Waren Sie schon einmal beim Neurologen oder Nervenarzt in Behandlung?</b>	Nein, auf dem Gebiet war bisher immer alles ok.

## Anhang 5: Musterlösung Patientenfälle

<b>Gruppe 1</b>		
<b>Wagner</b>	<b>Punkte</b>	
Verdachtsdiagnose	2	Lungenembolie bei Prostatakarzinom Lungenembolie bei Lymphom
	1	Lungenembolie Myelodysplastisches Syndrom
Begründung	10	<ul style="list-style-type: none"> <li>■ Vitalparameter: SpO<sub>2</sub> initial 82 %, 95 % unter 2l O<sub>2</sub></li> <li>■ EKG: Vorhofflimmern</li> <li>■ Labor: BSG erhöht, Panzytopenie</li> <li>■ B-Symptomatik: Gewichtsverlust, Nachtschweiß, Leistungsknick</li> <li>■ Alter des Patienten</li> <li>■ Blässe</li> <li>■ Infektanamnese: Pneumonie, Bronchitis</li> <li>■ Schwindel-/ Atemnotanamnese</li> <li>■ Makrohämaturie</li> <li>■ Synkope („schwarz vor Augen“)</li> </ul>
<b>Bircher</b>		
Verdachtsdiagnose	2	Herzinsuffizienz bei Tachyarrhythmia absoluta Herzinsuffizienz bei Vorhofflimmern
	1	Akute kardiale Dekompensation Akute Dekompensation einer Herzinsuffizienz Herzinsuffizienz Linksherzinsuffizienz Restriktion bei Pleuraerguss Instabile Angina Pectoris
Begründung	8	<ul style="list-style-type: none"> <li>■ Vitalparameter: Puls unregelmäßig, erhöht</li> <li>■ EKG: Vorhofflimmern, Tachykardie</li> <li>■ Untersuchungsbefund: Atemgeräusch abgeschwächt, RGs ubiquitär</li> </ul>

---

**Gruppe 1**


---

- Symptomatik der Herzinsuffizienz: Orthopnoe
  - vegetative Symptomatik: innere Unruhe, Zittern, Nervosität, Palpitationen, AP-Beschwerden
  - Alter der Patientin
  - Vorerkrankungen: arterielle Hypertonie, Diabetes Mellitus medikamentös therapiert
  - Harnwegsinfekt: Labor: Niereninsuffizienz, CRP erhöht; UStix positiv, Pollakisurie, Makrohämaturie, Nykturie
- 

**Lehner**

Verdachtsdiagnose	2	Hyperventilation bei Panikattacke Hyperventilation bei Panikstörung
	1	Panikattacke Panikstörung
Begründung	10	<ul style="list-style-type: none"> <li>▪ EKG: unauffällige Befunde</li> <li>▪ Labor: unauffällige Befunde</li> <li>▪ Symptomatik: Parästhesien an Händen und Beine</li> <li>▪ Alter des Patienten</li> <li>▪ situationsspezifischer Beginn der Symptomatik bekannt</li> <li>▪ Symptomatik ärztlich abgeklärt; keine somatischen Vorerkrankungen</li> <li>▪ Todesangst</li> <li>▪ Life-Event (Schwangerschaft der Ehefrau)</li> <li>▪ ängstliche Persönlichkeit</li> <li>▪ Alpträume, Schlafstörungen</li> </ul>

---



---

**Gruppe 2**


---

<b>Aimüller</b>	<b>Punkte</b>	
Verdachtsdiagnose	2	Lungenembolie bei Gerinnungsstörung Lungenembolie bei Antiphospholipidsyndrom Lungenembolie bei Hereditärer Thrombophilie
	1	Lungenembolie
Begründung	10	<ul style="list-style-type: none"> <li>▪ Vitalparameter: SpO<sub>2</sub> initial 92 %</li> <li>▪ EKG: Sinustachykardie, Zeichen der Rechtsherzbelastung (RSB), T-Negativierung V1-V2</li> <li>▪ Symptomatik: Tachykardie</li> <li>▪ Symptomatik: Tachypnoe</li> <li>▪ Symptomatik: atemabhängige Thoraxschmerzen mit trockenem Husten</li> <li>▪ Vorerkrankung: Gerinnungsstörung (familiär); Fehlgeburten</li> </ul>

---

---

**Gruppe 2**


---

- perakuter Beginn der Symptomatik
  - Schwindelanamnese
  - Bewusstlosigkeit (Synkope)
  - Panik, Todesangst
- 

**Albrecht**


---

Verdachtsdiagnose	2	Herzinsuffizienz bei Hypertropher obstruktiver Kardiomyopathie (HOCM)
		Herzinsuffizienz bei linksventrikulärer konzentrischer Hypertrophie
	1	Herzinsuffizienz
		Rechtsherzinsuffizienz
		Linksherzinsuffizienz
		Herzinsuffizienz bei Klappenitium
		Herzinsuffizienz bei DKMP
		Arrhythmogene rechtsventrikuläre Kardiomyopathie
		Herzinsuffizienz bei Z.n. Myokarditis
		Kardiomyopathie
		Hypertrophe Kardiomyopathie
Begründung	8	<ul style="list-style-type: none"> <li>▪ EKG: inkompletter LSB, Sinustachykardie</li> <li>▪ Untersuchungsbefund: neu aufgetretenes Herzgeräusch</li> <li>▪ Symptomatik der Herzinsuffizienz: Nykturie, Orthopnoe, Beinödeme</li> <li>▪ vegetative Symptomatik bei Belastung: Schwindel, Herzrasen, AP-Beschwerden, Dyspnoe</li> <li>▪ Alter des Patienten</li> <li>▪ Vorerkrankung: arterielle Hypertonie medikamentös therapiert</li> <li>▪ positive Familienanamnese: plötzlicher Herztod bei 2 Cousinen, Herzschrittmacher bei Vater</li> <li>▪ schleichender Beginn der Symptomatik (Monate)</li> </ul>

---


**Klein**


---

Verdachtsdiagnose	2	Ambulant erworbene Pneumonie (CAP)
		Bakterielle Pneumonie
	1	Pneumonie
		Akute Bronchitis
Begründung	5	<ul style="list-style-type: none"> <li>▪ Vitalparameter: Temperatur 39,8°C (Fieber, Schüttelfrost)</li> <li>▪ Labor: Leukozytose, CRP erhöht</li> <li>▪ grippale Symptomatik: Kopfschmerzen Gliederschmerzen</li> <li>▪ atemabhängige Schmerzen (Begleitpleuritis)</li> <li>▪ akuter Beginn der Symptomatik mit rapider Verschlechterung</li> </ul>

---

## Anhang 6: Anamnesebogen

Anamnesebogen 

Patient \_\_\_\_\_

PIN \_\_\_\_\_

### Hauptsymptome

---

---

---

---

---

### Medizinische Vorgeschichte

---

---

---

---

---

### Allergien und aktuelle Medikation

---

---

---

---

---

### Sozial- und Familienanamnese

---

---

---

---

---

### Systemübersicht

---

---

---

---

---

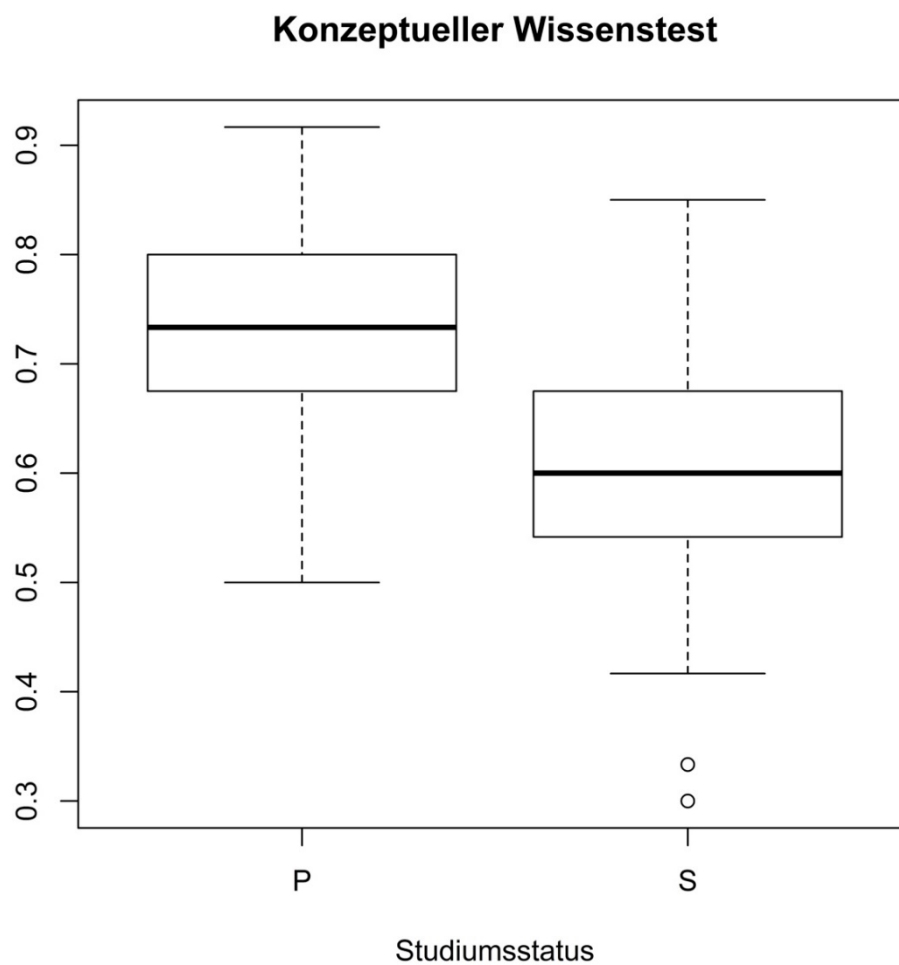
## Anhang 7: Überprüfung der Messinstrumente nach Studiumsstatus

Methodik: 2-seitiger t-Test zwischen den ungepaarten Gruppen S und P

$H_0$ : Die Differenz der arithmetischen Mittel ist gleich 0.

$H_1$ : Die Differenz der arithmetischen Mittel ist ungleich 0.

### 1. Konzeptueller Wissenstest

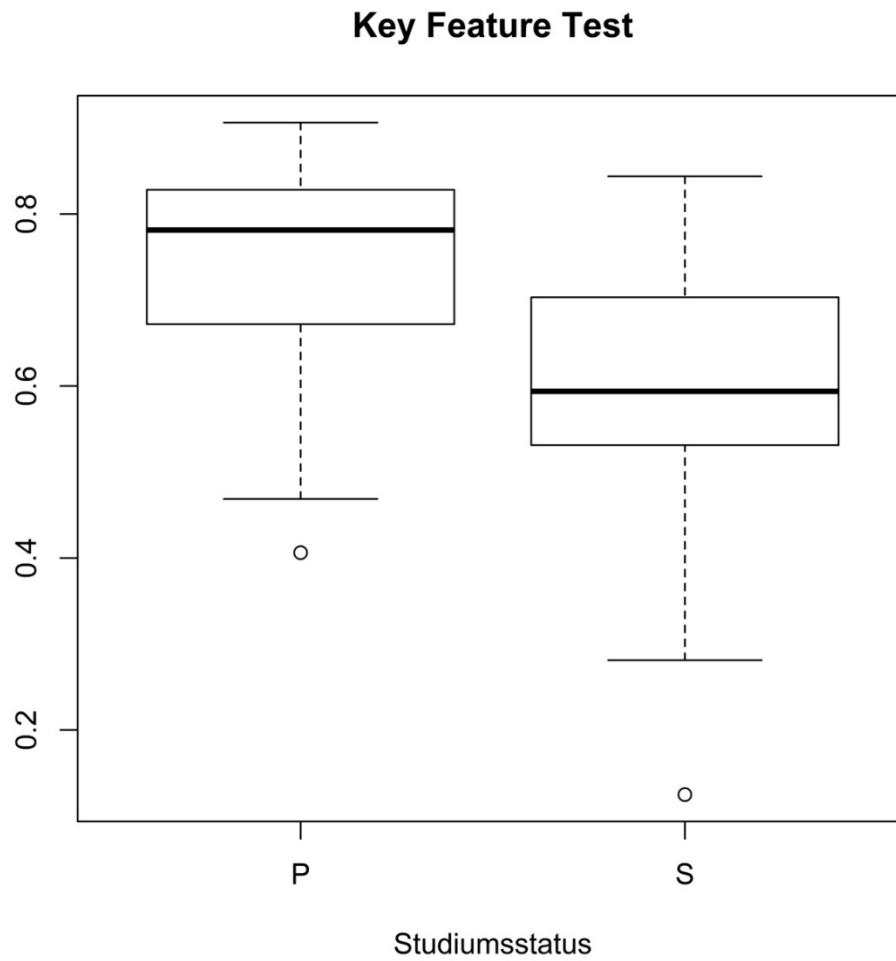


$t(81) = 5.14, p < .01$ ; 95 %-Konfidenzintervall [.08; .17]

$H_0$  kann zum Signifikanzniveau von 5 % abgelehnt werden.



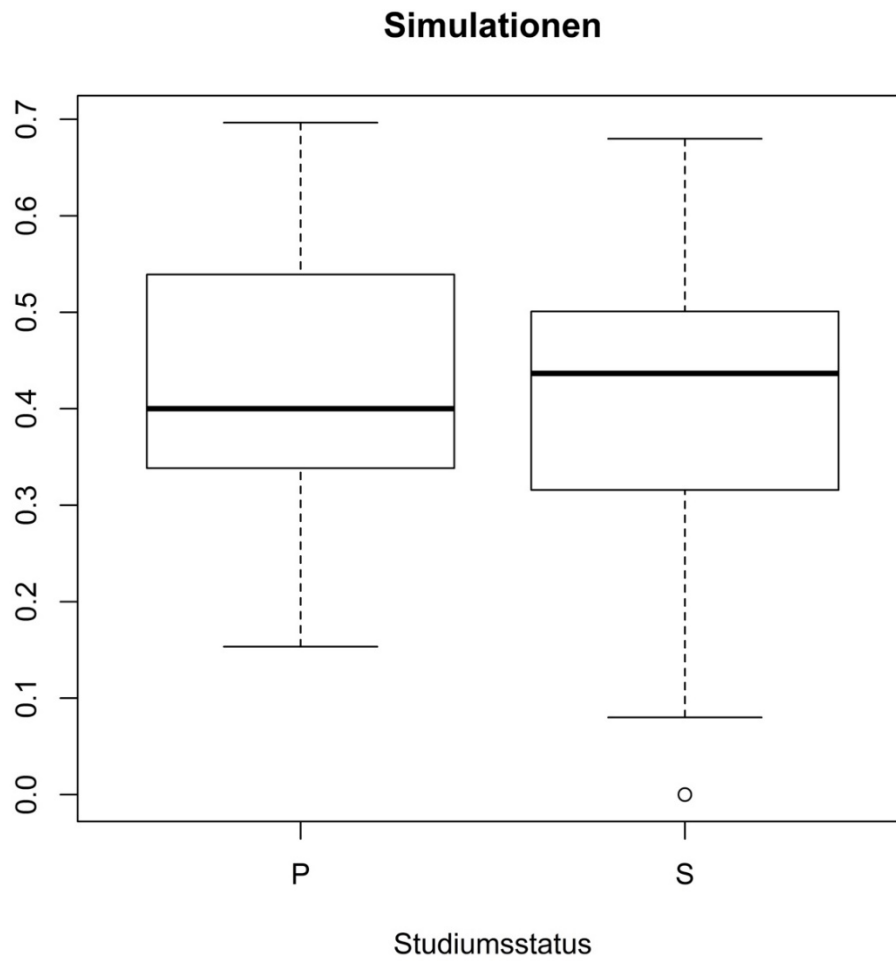
## 2. Key Feature Test



$t(84) = 4.82, p < .01$ ; 95 %-Konfidenzintervall [.08; .20]

$H_0$  kann zum Signifikanzniveau von 5 % abgelehnt werden.

### 3. Simulationen



$t(84) = 0.90, p = .37$ ; 95 %-Konfidenzintervall  $[-.08; .09]$

$H_0$  kann zum Signifikanzniveau von 5 % nicht abgelehnt werden.

## Anhang 8: Auswertungen der Forschungsfrage 3b) nach Studiumsstatus

Studierende im Klinischen Studienabschnitt:

(S)		Aimüller	Albrecht	Klein	Bircher	Lehner
	<b>Video/Start Live</b>	80.00	70.00	60.00	33.33	75.00
<b>Wahrscheinlichkeit für korrekte Diagnose (%)</b>	Video/Start Video	66.67	50.00	66.67	7.69	84.62
	<b>Live/Start Video</b>	79.92	84.62	84.62	58.33	50.00
	Live/Start Live	83.33	66.67	66.67	60.00	40.00
		2.00	2.33	0.75	6.00	0.55
<b>Odds Ratio</b> [95%-Konfidenzintervall]	<b>Video/Start Live</b>	[0.28, 14.20]	[0.40, 13.61]	[0.13, 4.29]	[0.56, 63.99]	[0.07, 4.01]
		0.67	2.75	2.75	0.93	1.50
	<b>Live/Start Video</b>	[0.09; 4.89]	[0.40; 18.88]	[0.40; 18.88]	[0.17; 5.15]	[0.27; 8.19]

Studierende im Praktischen Jahr:

(P)		Aimüller	Albrecht	Klein	Bircher	Lehner
	<b>Video/Start Live</b>	88.89	55.56	66.67	0.00	80.00
<b>Wahrscheinlichkeit für korrekte Diagnose (%)</b>	Video/Start Video	44.44	44.44	77.78	45.45	81.82
	<b>Live/Start Video</b>	81.82	90.91	90.91	33.33	66.67
	Live/Start Live	70.00	60.00	80.00	55.56	88.89
		10.00	1.56	0.57	0.00	0.89
<b>Odds Ratio</b> [95%-Konfidenzintervall]	<b>Video/Start Live</b>	[0.85, 117.02]	[0.24, 10.03]	[0.07, 4.64]	[0.00, ∞]	[0.10, 7.86]
		1.93	6.67	2.50	0.40	0.25
	<b>Live/Start Video</b>	[0.25; 14.89]	[0.60; 74.48]	[0.19; 32.79]	[0.06; 2.70]	[0.02; 3.04]

## 6 Verzeichnisse

### 6.1 Literaturverzeichnis

- [1] Nuland, S., *How We Die: Reflections on Life's Final Chapter*. New York: Vintage, 1994.
- [2] Seneca, L. A., "Liber VI, Epistula 57," in *Epistulae Morales ad Lucilium*, .
- [3] Berner, E. S. and Graber, M. L., "Overconfidence as a Cause of Diagnostic Error in Medicine," *Am. J. Med.*, vol. 121, no. 5, pp. S2–S23, 2008.
- [4] Graber, M. L., Franklin, N., and Gordon, R., "Diagnostic error in internal medicine," *Arch. Intern. Med.*, vol. 165, no. 13, pp. 1493–1499, 2005.
- [5] Schwartz, A. and Kostopoulou, O., "Clinical Reasoning in Medicine," in *Clinical Reasoning in the Health Professions*, 4th ed., J. Higgs, G. M. Jensen, S. Loftus, and N. Christensen, Eds. Edinburgh: Elsevier Health Sciences, 2019, pp. 223–234.
- [6] Kostopoulou, O., Mousoulis, C., and Delaney, B. C., "Information search and information distortion in the diagnosis of an ambiguous presentation," *Judgm. Decis. Mak.*, vol. 4, no. 5, pp. 408–418, 2009.
- [7] Graber, M. L., Wachter, R. M., and Cassel, C. K., "Bringing diagnosis into the quality and safety equations," *JAMA - J. Am. Med. Assoc.*, vol. 308, no. 12, pp. 1211–1212, 2012.
- [8] Keifenheim, K. E. *et al.*, "Teaching history taking to medical students: A systematic review," *BMC Med. Educ.*, vol. 15, no. 159, p. DOI 10.1186/s12909-015-0443-x, 2015.
- [9] Mayne, J., Weksel, W., and Sholtz, P., "Toward automating the medical history," *Mayo Clin. Proc.*, vol. 43, no. 1, pp. 1–25, 1968.
- [10] Alscher, M., "Computermedizin: Chancen für eine Individual Medizin," in *Die Zukunft der Individualmedizin*, 1st ed., R. Jütte, Ed. Köln: Deutscher Ärzte-Verlag GmbH, 2009, p. 136.
- [11] Bacher-Meck, S., "Erfahrungen mit elektronischen Anamnesen in einer nephrologischen Facharztpraxis," Eberhard Karls Universität zu Tübingen, 2015.
- [12] Berman, N. B., Durning, S. J., Fischer, M. R., Huwendiek, S., and Triola, M. M., "The role for virtual patients in the future of medical education," *Acad. Med.*, vol. 91, no. 9, pp. 1217–1222, 2016.
- [13] Dooley-Hash, S., "Educating Physicians: A Call for Reform of Medical School and Residency," *JAMA*, vol. 304, no. 11, pp. 1240–1241, 2010.
- [14] Graber, M. L., Wachter, R. M., and Cassel, C. K., "Bringing diagnosis into the quality and safety equations," *JAMA - Journal of the American Medical Association*. 2012.
- [15] Okuda, Y. *et al.*, "The utility of simulation in medical education: What is the evidence?," *Mt. Sinai J. Med.*, vol. 76, no. 4, pp. 330–343, 2009.
- [16] Mamede, S., Van Gog, T., Sampaio, A. M., De Faria, R. M. D., Maria, J. P., and Schmidt, H. G., "How can students' diagnostic competence benefit most from practice with clinical cases? the effects of structured reflection on future diagnosis of the same and novel diseases," *Acad. Med.*, vol. 89, no. 1, pp. 121–127, 2014.

- [17] Boshuizen, H. P. and Schmidt, H. G., "The Development of Clinical Reasoning Expertise," in *Clinical Reasoning in the Health Professions*, 4th ed., J. Higgs, G. Jensen, S. Loftus, and N. Christensen, Eds. Edinburgh: Elsevier Health Sciences, 2019, pp. 57–66.
- [18] Jones, M., Edwards, I., and Jensen, G. M., "Clinical Reasoning in Physiotherapy," in *Clinical Reasoning in the Health Professions*, 4th ed., J. Higgs, G. M. Jensen, S. Loftus, and N. Christensen, Eds. Edinburgh: Elsevier Health Sciences, 2019, pp. 247–260.
- [19] Kopp, V., Stark, R., and Fischer, M. R., "Förderung von Diagnosekompetenz in der medizinischen Ausbildung durch Implementation eines Ansatzes zum fallbasierten Lernen aus Lösungsbeispielen," *GMS Z Med Ausbild*, vol. 24, no. 2, p. Doc 107, 2007.
- [20] Mandl, H., Kopp, B., and Dvorak, S., *Aktuelle theoretische Ansätze und empirische Befunde im Bereich der Lehr-Lern-Forschung – Schwerpunkt Erwachsenenbildung* –. 2004.
- [21] Jensen, G. M., Resnik, L. J., and Haddad, A. M., "Expertise and Clinical Reasoning," in *Clinical Reasoning in the Health Professions*, 4th ed., J. Higgs, G. M. Jensen, S. Loftus, and N. Christensen, Eds. Edinburgh: Elsevier Health Sciences, 2019, pp. 67–76.
- [22] Christensen, N., Villanueva, C., and Grieve, S., "Learning Reasoning using Simulation," in *Clinical Reasoning in the Health Professions*, 4th ed., J. Higgs, G. M. Jensen, S. Loftus, and N. Christensen, Eds. Edinburgh: Elsevier Health Sciences, 2019, pp. 455–464.
- [23] Bundesministerium der Justiz und für Verbraucherschutz, "Approbationsordnung für Ärzte," 2019. [Online]. Available: [https://www.gesetze-im-internet.de/\\_appro\\_2002/BJNR240500002.html](https://www.gesetze-im-internet.de/_appro_2002/BJNR240500002.html). [Accessed: 29-Jan-2019].
- [24] Richter-Kuhlmann, E., "Ärztliche Kunst lernen und lehren," *Dtsch. Arztebl.*, vol. 116, no. 3, pp. C62–C63, 2019.
- [25] Technische Universität München, "Studienordnung für den Studiengang Medizin der Technischen Universität München Präambel," München, 2011.
- [26] Kiesewetter, J., Kager, M., Lux, R., Zwissler, B., Fischer, M. R., and Dietz, I., "German undergraduate medical students' attitudes and needs regarding medical errors and patient safety-A national survey in Germany," *Med. Teach.*, vol. 36, no. 6, pp. 505–510, 2014.
- [27] Cook, D. A. *et al.*, "Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis," *Med. Teach.*, vol. 35, no. 1, pp. e867–e898, 2013.
- [28] Schmidmaier, R. *et al.*, "Learning the facts in medical school is not enough: Which factors predict successful application of procedural knowledge in a laboratory setting?," *BMC Med. Educ.*, vol. 13, no. 28, p. doi: 10.1186/1472-6920-13-28, 2013.
- [29] Ilgen, J. S. *et al.*, "Assessing diagnostic reasoning: A consensus statement summarizing theory, practice, and future needs," *Acad. Emerg. Med.*, vol. 19, no. 2, pp. 1454–1461, 2012.
- [30] Heitzmann, N. *et al.*, "Facilitating Diagnostic Competences in Simulations: A Conceptual Framework and a Research Agenda for Medical and Teacher Education," (accepted, revisions pending) *Frontline Learning Research*.

- [31] Smith, M. and Higgs, J., “Learning about Factors influencing Clinical Decision Making,” in *Clinical Reasoning in the Health Professions*, 4th ed., J. Higgs, G. M. Jensen, S. Loftus, and N. Christensen, Eds. Edinburgh: Elsevier Health Sciences, 2019, pp. 445–454.
- [32] Wisniewski, E. J. and Medin, D. L., “On the interaction of theory and data in concept learning,” *Cogn. Sci.*, vol. 18, no. 2, pp. 221–281, 1994.
- [33] Staatsinstitut für Schulqualität und Bildungsforschung München, “Pädagogisch diagnostizieren im Schulalltag.” [Online]. Available: [http://www.isb.bayern.de/download/7409/paedagogisch\\_diagnostizieren.pdf](http://www.isb.bayern.de/download/7409/paedagogisch_diagnostizieren.pdf). [Accessed: 09-Jan-2019].
- [34] Baumert, J. and Kunter, M., “Das Kompetenzmodell von COACTIV,” in *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*, 1st ed., M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, and M. Neubrand, Eds. Münster: Waxmann, 2011, p. 369.
- [35] Weinert, F. E., “Vergleichende Leistungsmessung in der Schule - eine umstrittene Schulverständlichkeit,” in *Leistungsmessungen in Schulen*, F. E. Weinert, Ed. Weinheim: Beltz, 2001, pp. 17–31.
- [36] Chomsky, N., *Regeln und Repräsentationen*. Frankfurt: Suhrkamp, 1981.
- [37] Koeppen, K., Hartig, J., Klieme, E., and Leutner, D., “Current Issues in Competence Modeling and Assessment,” *Zeitschrift für Psychol. / J. Psychol.*, vol. 216, no. 2, pp. 61–73, 2008.
- [38] Schmidt, H. G. and Rikers, R. M. J. P., “How expertise develops in medicine: knowledge encapsulation and illness script formation,” *Med. Educ.*, vol. 41, no. 12, pp. 1133–1139, 2007.
- [39] Boshuizen, H. P. A. and Schmidt, H. G., “On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices,” *Cogn. Sci.*, vol. 16, no. 2, pp. 153–184, 1992.
- [40] Higgs, J. and Jensen, G. M., “Clinical Reasoning: Challenges of Interpretation and Practice in the 21st Century,” in *Clinical Reasoning in the Health Professions*, 4th ed., J. Higgs, G. M. Jensen, S. Loftus, and N. Christensen, Eds. Edinburgh: Elsevier Health Sciences, 2019, pp. 3–12.
- [41] Borleffs, J. C. C., Custers, E. J. F. M., Van Gijn, J., and Ten Cate Olle, T. J., “‘Clinical reasoning theater’: A new approach to clinical reasoning education,” *Acad. Med.*, vol. 78, no. 3, pp. 322–325, 2003.
- [42] Hardin, L., “Research in Medical Problem Solving: A Review,” *J. Vet. Med. Educ.*, vol. 30, no. 3, pp. 230–235, 2003.
- [43] Trowbridge, R. L., Rencic, J. J., and Durning, S. J., *Teaching Clinical Reasoning*, 1st ed. Philadelphia: American College of Physicians, 2015.
- [44] Barrows, H. S. and Pickel, G. C., *Developing Clinical Problem-Solving Skills: A Guide To More Effective Diagnosis And Treatment*, 1st ed. New York: Norton Medical Books, 1991.
- [45] Gabler Wirtschaftslexikon, “Wissen.” [Online]. Available: <https://wirtschaftslexikon.gabler.de/definition/wissen-47196/version-270462>. [Accessed: 29-Jan-2019].
- [46] Mayer, R. E., “Applying the science of learning to medical education,” *Med. Educ.*,

- vol. 44, no. 6, pp. 543–549, 2010.
- [47] Paris, S. G. and Wixson, K. K., “Becoming a Strategic Reader,” *Contemp. Educ. Psychol.*, vol. 316, no. 8, pp. 293–316, 1983.
- [48] Stark, R., Kopp, V., and Fischer, M. R., “Case-based learning with worked examples in complex domains: Two experimental studies in undergraduate medical education,” *Learn. Instr.*, vol. 21, no. 1, pp. 22–33, 2011.
- [49] Jones, M., Jensen, G. M., and Edwards, I., “Clinical reasoning in physiotherapy,” in *Clinical Reasoning in the Health Professions*, 2nd ed., J. Higgs and M. Jones, Eds. Oxford: Butterworth-Heinemann, 2000, pp. 117–127.
- [50] Klemme, B. and Siegmann, G., “Zentrale Elemente und Strategien des Clinical Reasoning,” in *Clinical Reasoning: Therapeutische Denkprozesse lernen*, 2nd ed., B. Klemme and G. Siegmann, Eds. Stuttgart: Georg Thieme Verlag, 2015, pp. 28–48.
- [51] Rikers, R. M. J. P., Loyens, S. M. M., and Schmidt, H. G., “The role of encapsulated knowledge in clinical case representations of medical students and family doctors,” *Med. Educ.*, vol. 38, no. 10, pp. 1035–1043, 2004.
- [52] Patel, V. and Groen, G., “The general and specific nature of medical expertise: A critical look,” in *Toward a general theory of expertise*, K. Ericsson and J. Smith, Eds. Cambridge: Cambridge University Press, 1991, pp. 93–125.
- [53] GRANT, J. and MARSDEN, P., “Primary knowledge, medical education and consultant expertise,” *Med. Educ.*, vol. 22, no. 3, pp. 173–179, 1988.
- [54] Lubarsky, S., Dory, V., Audétat, M.-C., Custers, E., and Charlin, B., “Using script theory to cultivate illness script formation and clinical reasoning in health professions education,” *Can. Med. Educ. J.*, vol. 6, no. 2, pp. e61–e70, 2015.
- [55] Rikers, R. M. J. P., Schmidt, H. G., and Moulaert, V., “Biomedical knowledge: Encapsulated or two worlds apart?,” *Appl. Cogn. Psychol.*, vol. 19, no. 2, pp. 223–231, 2005.
- [56] Pätzold, G., “Berufliche Handlungskompetenz,” in *Wörterbuch Berufs- und Wirtschaftspädagogik*, F. Kaiser and G. Pätzold, Eds. Julius Klinkhardt, 1999, pp. 57–58.
- [57] Durning, S. J., Dong, T., Artino, A. R., van der Vleuten, C., Holmboe, E., and Schuwirth, L., “Dual processing theory and experts’ reasoning: exploring thinking on national multiple-choice questions,” *Perspect. Med. Educ.*, vol. 4, no. 4, pp. 168–175, 2015.
- [58] Mattingly, C. and Fleming, M. H., *Clinical Reasoning: Forms of Inquiry in a Therapeutic Practice*. Philadelphia: F.A. Davies, 1994.
- [59] Higgs, J. and Jones, M., “Multiple Spaces of Choice, Engagement and Influence in Clinical Decision Making,” in *Clinical Reasoning in the Health Professions*, J. Higgs, G. M. Jensen, S. Loftus, and Ni. Christensen, Eds. Edinburgh: Elsevier Health Sciences, 2019, pp. 33–43.
- [60] Paul, R. and Elder, L., *A Guide for Educators to Critical Thinking Competency Standards*. Dillon Beach, CA, 2007.
- [61] Handgraaf, M., Klemme, B., and Nauerth, A., “Entwicklung eines Schulungskonzepts zum ‘Clinical Reasoning’ in therapeutischen Berufen,” Bielefeld, 2004.
- [62] Kahneman, D., *Thinking, Fast and Slow*. New York, 2011.

- [63] Jones, M., "Clinical Reasoning: Fundament der klinischen Praxis und Brücke zwischen den Ansätzen der Manuellen Therapie," *Man. Ther.*, no. 4, pp. 3–9, 1997.
- [64] Marcum, J. A., "An integrated model of clinical reasoning: Dual-process theory of cognition and metacognition," *J. Eval. Clin. Pract.*, vol. 18, no. 5, pp. 954–961, 2012.
- [65] Seel, N. M., *Psychologie des Lernens*. München: Ernst Reinhardt, 2000.
- [66] Wainwright, S. F., Shepard, K. F., Harman, L. B., and Stephens, J., "Novice and Experienced Physical Therapist Clinicians: A Comparison of How Reflection Is Used to Inform the Clinical Decision-Making Process," *Phys. Ther.*, vol. 90, no. 1, pp. 75–88, 2010.
- [67] Quintana, C. *et al.*, "A scaffolding design framework for software to support science inquiry," *J. Learn. Sci.*, vol. 13, no. 3, pp. 337–386, 2004.
- [68] Sweller, J., Van Merriënboer, J. J. G., and Paas, F. G. W. C., "Cognitive Architecture and Instructional Design," *Educ. Psychol. Rev.*, vol. 10, no. 3, pp. 251–296, 1998.
- [69] Cook, D. A., "How much evidence does it take? A cumulative meta-analysis of outcomes of simulation-based education," *Med. Educ.*, vol. 48, no. 8, pp. 750–760, 2014.
- [70] Yardley, S., Teunissen, P. W., and Dornan, T., "Experiential learning: AMEE Guide No. 63," *Med. Teach.*, vol. 34, pp. e102–e115, 2012.
- [71] De Jong, T., "Instruction based on computer simulations," in *Handbook of research on learning and instruction*, R. Mayer and P. Alexander, Eds. New York: Routledge, 2011, pp. 446–466.
- [72] Peeraer, G. *et al.*, "Clinical skills training in a skills lab compared with skills training in internships: Comparison of skills development curricula," *Educ. Heal. Chang. Learn. Pract.*, vol. 20, no. 3, p. 125, 2007.
- [73] Rudolph, J. W., Raemer, D. B., and Simon, R., "Establishing a safe container for learning in simulation: The role of the Presimulation Briefing," *Simul. Healthc.*, vol. 9, no. 6, pp. 339–349, 2014.
- [74] Fischer, M. R., "CASUS - An authoring and learning tool supporting diagnostic reasoning," *Zeitschrift für Hochschuldidaktik*, vol. 1, no. 1, pp. 87–98, 2000.
- [75] Cook, D. A., Erwin, P. J., and Triola, M. M., "Computerized Virtual Patients in Health Professions Education: A Systematic Review and Meta-Analysis," *Acad. Med.*, vol. 85, no. 10, pp. 589–602, 2010.
- [76] Blömeke, S., Gustafsson, J. E., and Shavelson, R. J., "Beyond dichotomies: Competence viewed as a continuum," *Zeitschrift für Psychol. / J. Psychol.*, vol. 223, no. 1, pp. 3–13, 2015.
- [77] Miller, G. E., "The assessment of clinical skills/competence/performance," *Acad. Med.*, vol. 65, no. 9, pp. 63–67, 1990.
- [78] Kopp, V., Möltner, A., and Fischer, M. R., "Key-Feature-Probleme zum Prüfen von prozeduralem Wissen: Ein Praxisleitfaden," *GMS Z. Med. Ausbild.*, vol. 23, no. 3, p. Doc50, 2006.
- [79] Hrynchak, P., Glover Takahashi, S., and Nayer, M., "Key-feature questions for assessment of clinical reasoning: A literature review," *Med. Educ.*, vol. 48, pp. 870–883, 2014.
- [80] Page, G., Bordage, G., and Allen, T., "Developing key-feature problems and



- examinations to assess clinical decision-making skills,” *Acad. Med.*, vol. 70, no. 2, pp. 104–110, 1995.
- [81] Schuwirth, L. W. T., Durning, S. J., Norman, G. R., and van der Vleuten, C. P. M., “Assessing Clinical Reasoning,” in *Clinical Reasoning in the Health Professions*, 4th ed., J. Higgs, G. M. Jensen, S. Loftus, and N. Christensen, Eds. Edinburgh: Elsevier Health Sciences, 2019, pp. 407–415.
- [82] Norman, G. R. and Schmidt, H. G., “The psychological basis of problem-based learning: A review of the evidence,” *Acad. Med.*, vol. 67, no. 9, pp. 557–565, 1992.
- [83] Gruppen, L., “Clinical Reasoning: Defining It, Teaching It, Assessing It, Studying It,” *West. J. Emerg. Med.*, vol. 18, no. 1, pp. 4–7, 2017.
- [84] Schuwirth, L. W. T., Verheggen, M. M., Van Der Vleuten, C. P. M., Boshuizen, H. P. A., and Dinant, G. J., “Do short cases elicit different thinking processes than factual knowledge questions do?,” *Med. Educ.*, vol. 35, no. 4, pp. 348–356, 2001.
- [85] Park, W. B., Kang, S. H., and Lee, Y., “Reflect the Clinical Reasoning Ability of Medical Students?,” *Am. Med. J.*, vol. 350, no. 1, pp. 64–67, 2015.
- [86] Chenot, J.-F. and Ehrhardt, M., “Objective structured clinical examination (OSCE) in medical training: An alternative to a written exam | Objective structured clinical examination (OSCE) in der medizinischen Ausbildung: Eine alternative zur Klausur,” *Z. Allgemeinmed.*, vol. 79, no. 9, pp. 437–442, 2003.
- [87] Petrusa, E. R., “Clinical Performance Assessments,” in *International Handbook of Research in Medical Education*, G. R. Norman, C. van der Vleuten, and D. I. Newble, Eds. Dordrecht: Springer Netherlands, 2002, pp. 673–709.
- [88] Clauser, B. E. and Schuwirth, L., “The Use of Computers in Assessment,” in *International Handbook of Research in Medical Education*, G. R. Norman, C. van der Vleuten, and D. R. Newble, Eds. Dordrecht: Springer Netherlands, 2002, pp. 757–792.
- [89] Berliner, D., Schneider, N., Welte, T., and Bauersachs, J., “The differential diagnosis of dyspnoea,” *Dtsch. Arztebl. Int.*, vol. 113, no. 49, pp. 834–844, 2016.
- [90] Cuschieri, A., Gleeson, F. A., Harden, R. M., and Wood, R. A. B., “A new approach to a final examination in surgery. Use of the objective clinical examinations,” *Ann. R. Coll. Surg. Engl.*, vol. 61, no. 5, pp. 400–405, 1979.
- [91] Fischer, M. R., Kopp, V., Holzer, M., Ruderich, F., and Jünger, J., “A modified electronic key feature examination for undergraduate medical students: Validation threats and opportunities,” *Med. Teach.*, vol. 27, no. 5, pp. 450–455, 2005.
- [92] Rotthoff, T. *et al.*, “Comparison between Long-Menu and Open-Ended Questions in computerized medical assessments. A randomized controlled trial,” *BMC Med. Educ.*, vol. 6, no. 50, 2006.
- [93] Bauer, D., Holzer, M., Kopp, V., and Fischer, M. R., “Pick-N multiple choice-exams: A comparison of scoring algorithms,” *Adv. Heal. Sci. Educ.*, p. DOI 10.1007/s10459-010-9256-1, 2011.
- [94] Tabachnick, B. G. and Fidell, L. S., *Using Multivariate Statistics*, 6th ed. Edinburgh: Pearson Education Limited, 2013.
- [95] Office of Educational Assessment, “Understanding Item Analyses,” *University of Washington*, 2019. [Online]. Available: <http://www.washington.edu/assessment/scanning-scoring/scoring/reports/item-analysis/>. [Accessed: 04-Feb-2019].

- [96] Bortz, J. and Doering, N., *Forschungsmethoden und Evaluation*, 3rd ed. Berlin: Springer, 2002.
- [97] Bortz, J. and Doering, N., *Forschungsmethoden und Evaluation für Human und Sozialwissenschaftler*, 4th ed. Berlin: Springer, 2006.
- [98] DFG-Forschergruppe 2358, “Teilprojekt 6.” [Online]. Available: <https://www.for2385.uni-muenchen.de/teilprojekte1/teilprojekt6/index.html>. [Accessed: 06-May-2019].
- [99] Krause, U.-M., *Feedback und kooperatives Lernen*, 1st ed. Münster: Waxmann, 2007.
- [100] Wimmers, P. F., Schmidt, H. G., and Splinter, T. A. W., “Influence of clerkship experiences on clinical competence,” *Med. Educ.*, vol. 40, pp. 450–458, 2006.

## 6.2 Tabellenverzeichnis

Tabelle 1	Kennzeichen verschiedener Expertisestadien im Clinical Reasoning Prozess .....	11
Tabelle 2	Ablauf der Studie .....	20
Tabelle 3	Demographische Angaben des ausgewerteten Studienkollektivs .....	21
Tabelle 4	Diagnosen der Patientenfälle aus den Simulationen .....	25
Tabelle 5	Punktesystem zur Auswertung der Diagnosen der Patientenfälle aus den Simulationen .....	27
Tabelle 6	Aufteilung der Testergebnisse aus den Simulationen nach Gruppe, Startbedingung und Studiumsstatus .....	29
Tabelle 7	Schwierigkeit und Trennschärfe der Frageitems im Konzeptuellen Wissenstest.....	31
Tabelle 8	Schwierigkeit und Trennschärfe der Fallitems im Key Feature Test.....	33
Tabelle 9	Schwierigkeit und Trennschärfe der Frageitems im Key Feature Test.....	34
Tabelle 10	Auswertung der Testergebnisse im Konzeptuellen Wissenstest.....	36
Tabelle 11	Auswertung der Testergebnisse im Key Feature Test .....	36
Tabelle 12	Auswertung der Testergebnisse in den Simulationen .....	37
Tabelle 13	Auswertung der Testergebnisse im Fall Aimüller .....	39
Tabelle 14	Auswertung der Testergebnisse im Fall Albrecht.....	41
Tabelle 15	Auswertung der Testergebnisse im Fall Klein.....	43
Tabelle 16	Auswertung der Testergebnisse im Fall Bircher.....	45
Tabelle 17	Auswertung der Testergebnisse im Fall Lehner .....	47
Tabelle 18	Korrelation nach Pearson zwischen der Performanz in Wissenstests und Simulationen insgesamt und unter Aufteilung nach Gruppe .....	49
Tabelle 19	Gruppe 1: Korrelation nach Pearson zwischen der Performanz in Wissenstests und Simulationen insgesamt und unter Aufteilung nach Simulationsart .....	50
Tabelle 20	Gruppe 2: Korrelation nach Pearson zwischen der Performanz in Wissenstests und Simulationen insgesamt und unter Aufteilung nach Simulationsart .....	51
Tabelle 21	Wahrscheinlichkeiten für korrekte Diagnosen in den Patientenfällen der Simulationen unter Aufteilung nach Simulationsart.....	52

---

Tabelle 22	Effekt der Reihenfolge auf die Wahrscheinlichkeiten für korrekte Diagnosen in den Patientenfällen der Simulationen unter Aufteilung nach Simulationsart und Startbedingung .....	53
Tabelle 23	Performanz in den Patientenfällen der Simulationen unter Aufteilung nach Gruppe und Startbedingung.....	53

### 6.3 Abbildungsverzeichnis

Abbildung 1.	Einflussfaktoren und Auswirkungen von Diagnosekompetenzen .....	13
Abbildung 2.	Arbeitsmodell für Diagnosekompetenzen der DFG-Forschergruppe COSIMA .....	15
Abbildung 3.	Modifizierte Wissenspyramide nach Miller und Messinstrumente zur Erfassung einzelner Teilbereiche von Diagnosekompetenzen.....	17
Abbildung 4.	Schwierigkeit und Trennschärfe der ausgewerteten Frageitems (n = 30) im Konzeptuellen Wissenstest .....	32
Abbildung 5.	Schwierigkeit und Trennschärfe der Fallitems im Key Feature Test.....	33
Abbildung 6.	Schwierigkeit und Trennschärfe der ausgewerteten Frageitems (n = 32) im Key Feature Test.....	35
Abbildung 7.	Verteilung der Testergebnisse der Messinstrumente .....	35
Abbildung 8.	Korrelation nach Pearson zwischen der Performanz im Konzeptuellen Wissenstests und Key Feature Test unter Berücksichtigung des Studiumsstatus .....	48
Abbildung 9.	Korrelation nach Pearson zwischen der Performanz in Konzeptuellem Wissenstest und Simulationen sowie Key Feature Test und Simulationen unter Berücksichtigung des Studiumsstatus .....	49

## 7 Danksagung

Ich danke Herrn Professor Fischer für die Überlassung des Themas sowie die intensive, geduldige und stets sehr freundliche Betreuung über den ganzen Zeitraum der Anfertigung der Arbeit.

Frau Dr. Hannah Gerstenkorn danke ich herzlich für die engagierte Hilfe und konstruktive Kritik im Verlauf.

Herrn Maximilian Fink danke ich für die Rekrutierung in das Forschungsprojekt, die gemeinsame Datenerhebung sowie die umfangreiche, positive Unterstützung.

Vielen Dank allen Mitarbeitern und Doktoranden an den Lehrstühlen für Didaktik und Ausbildungsforschung in der Medizin, für Statistik und für Health Services Management der LMU München für jegliche Form der freundlichen und bedingungslosen Zusammenarbeit.

Eine große Honorierung gilt allen engagierten und kooperativen Studienteilnehmenden für die Messung ihrer klinischen Diagnosekompetenzen in einer simulierten Lernumgebung.

Ich danke meinen Eltern. Meiner Familie und meinen Freunden danke ich für die bedingungslose Unterstützung, Begleitung und Loyalität.

Gewidmet ist dieses Schriftstück allen Lebenden und Toten, die meinen Lebensweg bereichern.

## 8 Eidesstattliche Versicherung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel Messung von klinischen Diagnosekompetenzen bei Medizinstudierenden in einer simulierten Lernumgebung selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, den 07. Juni 2020

Victoria Reitmeier